2008

Development of a structure-based drug design module for a bioengineering lab

Aarthi Janakiraman
San Jose State University

DEVELOPMENT OF A STRUCTURE-BASED DRUG DESIGN MODULE FOR A BIOENGINEERING LAB

A Thesis

Presented to

The Faculty of Department of Chemical and Materials Engineering

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Aarthi Janakiraman

August 2008

APPROVED FOR THE DEPARTMENT OF CHEMICAL AND MATERIALS ENGINEERING

Melanie McNeil 04-07-08

Dr. Melanie McNeil

Dr. Brooke Lustig

Peter Nemes

Dr. Peter Nemes

APPROVED FOR THE UNIVERSITY

David K Bruck

ABSTRACT

DEVELOPMENT OF STRUCTURE-BASED DRUG DESIGN MODULE FOR BIOENGINEERING LAB

by

Aarthi Janakiraman

This thesis is focused on the development of a laboratory module for structure-based drug design. The process of structure-based drug design is thoroughly explored and simplified into five steps involving choosing the target, visualizing the target structure, identifying the binding site, docking the ligands, and evaluating them. The drug design for osteoarthritis was performed using the COX-2 enzyme as the drug target. Five ligand compounds, acetyl salicylic acid, rofecoxib, celecoxib, SC-558, and fucoxanthin were docked to the drug target individually with the help of UCSF DOCK software. Three ligands (rofecoxib, celecoxib, SC-558) unambiguously showed selective affinity for COX-2. The docked ligands, when evaluated using Lipinski's analysis, showed less absorption capabilities with increase in molecular weight and partition coefficient.

ACKNOWLEDGEMENT

I would like to thank my reading committee for their valuable suggestions and guidance throughout this research work. I am deeply thankful to Dr. Melanie McNeil not only for the guidance, but also for the encouragement and inspiration during the thesis work. I am grateful to Dr. Brooke Lustig for his insights into molecular modeling and valuable advice that made the science interesting for me. I am thankful to Dr. Peter Nemes for his inquisitive questions that shaped this thesis. Lastly, I am indebted to my husband Ramkumar and my daughter Kriya for their unwavering support and patience. I also thank my parents, my sister, and all my friends, especially Ms.Vrushali Tembe, Ms. Bagya Sodem, and Ms. Usha Jayanandam who were concerned and constantly encouraged me to complete this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FILES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

Modern medicine has made tremendous leaps in the field of drug discovery. The drug design industry is now one of the major players in the bioinformatics and biotechnology industries. Many pharmaceutical companies have a biotech and a bioinformatics unit in their research departments. The past two decades have seen the development of numerous procedures to diagnose and treat patients. The objective of this thesis proposal is to prepare a laboratory module for drug design using bioinformatics tools.

1.1 Economic Significance

Drug research is comparatively less expensive than health care. The National Institute of General Medical Sciences (NIGMS) states that in 1990, research and developmental costs were only 3.7 % of health care costs [1]. Trying to cure a disease with insufficient research leads to unsuccessful treatment and soaring health care costs resulting in severe economic impact. Research is the step which reduces health care costs while potentially saving lives.

Drug discovery is a time consuming process involving a lot of money. The process of discovering a drug and getting it approved by the FDA costs a pharmaceutical biotech company millions of dollars. In the year 1997, the top twenty pharmaceutical companies spent $16 billion on drug development alone [2]. Thousands of compounds were screened and studied in order to find one drug. On a positive note, when a drug is approved, the company gets billions of dollars in revenue and no other company can

duplicate it, as it is patented for at least 20 years. For example, in 2000, Prilosec, a drug commonly used to treat stomach ulcers, earned $4.102 billion in sales [2].

1.2 Time Factor

The development of any prospective drug begins with years of strenuous research to figure out the complexity of the medical problem. Often, it takes 5 years, minimum, to discover the drug, 2-5 years of preclinical testing, and 3-10 years of clinical testing. A patent is granted for 20 years at the end of the preclinical phase. The whole process takes 5 -10 years after which the drug is forwarded to the FDA (Food and Drug Administration) for final approval [3]. This means that the company is left with less than 10 years to earn back the research and development cost before the patent expires.

1.3 Drug Design

The development of a new drug is called drug discovery. Drug design is the first step in drug discovery research. This is a process in which drugs are invented based on the disease. Drugs are usually designed in a manner to bind to, interact with, and affect the activity of the key molecule under study.

Generally a pharmaceutical company aims to design a drug based on certain theoretical concepts but the effectiveness of this aspect is far from optimal. Drugs have been discovered accidentally, like penicillin. Another example of an accidentally discovered drug is Cisplatin [4]. The drug was discovered while a NIGMS (National Institute of General Medical Sciences) grantee was studying the effect of electric fields on bacteria. The researcher noticed that the platinum electrodes used in his experiment were inhibiting the bacteria from dividing. The platinum compounds had anti-tumor

effects and now Cisplatin is the leading drug used in treatment for testicular, ovarian and bladder cancers. Not all drugs have been found accidentally; most drugs are found after years of intensive research and hard work.

1.4 Types of Drug Design

There are two main types of research in drug design. The first method, called High Throughput Screening (HTS), identifies active compounds in collections or libraries as quickly as possible and with high statistical accuracy. Success in HTS depends on uncontrollable factors. For instance, HTS hits (successfully bound compounds) eventually decompose and some HTS hits have poor physico-chemical properties for use as potential drugs. Many compounds on the order of thousands or more must be tested to find a hit.

The other method, structure-based drug design, involves detailed knowledge of the binding sites of targets (such as proteins) associated with the disease. A drug's effectiveness depends on structural interaction with the receptor or target molecule. The most common model is the one in which the drug molecule fits itself into the crevice of the target protein similar to a key in a lock. This strategy results in inhibition of the protein's function, and ultimately halts the progress of the disease.

Structure-based drug design offers tremendous potential and a powerful alternative to traditional screening techniques. This drug design method involves basic knowledge of bioinformatics, proteomics, biochemistry, and computer modeling of 3-dimensional protein structures.

1.5 AIDS Drug Design through Structure-based Approach

Since many drugs have been discovered and developed in the past for treatment of AIDS (Acquired Immuno Deficiency Syndrome), it serves as an example for structure-based drug design. In the 1980s, the world of medicine witnessed the AIDS disaster. Eventually, doctors realized that a killer virus was infecting and destroying the immune system of the body.

In the late 1990s, deaths due to AIDS were drastically reduced when new drugs were introduced. Scientists determined the structure of HIV (Human Immunodeficiency Virus) protease and recognized that by inhibiting the enzyme, viral multiplication could be prevented. Researchers analyzed the HIV protease and developed drugs called Protease Inhibitors. These inhibitors prevented viral replication and prolonged the life span of AIDS patients.

In structure-based drug design, the concept is to map every point in the protein molecule and deduce ways to stop the protein from working. Funded by the National Institute of General Medical Sciences (NIGMS), three drugs (Norvir, Fortovase, and Viracept) took eight years to come into the market through fast tracking by the Food and Drug Administration (FDA). Other HIV drugs can be found on the NIGMS website. Apart from those intended to be used for the treatment of AIDS, many other drugs were designed based on structure for diseases like glaucoma, arthritis, flu, breast cancer, and trypanosoma.

The intention is to prepare a simple and a basic module on structure-based drug design for basic understanding of the drug design process for a bioinformatics graduate

course. This module will create an appreciation of the process of drug design and of how small molecules bind to the target macromolecule.

CHAPTER TWO

BACKGROUND

2.1 Process of Structure-based Drug Design

Structure-based drug design plays a major role in the development of novel drugs against many diseases. Recent advances in large scale determination of protein structures are improving the drug discovery process by starting with the protein structure and using it to design and identify new ligands. The objective of the background is to cover the significant steps involved in structure-based drug design.

2.2 Steps Involved in Structure-based Drug Design

1. Identification of drug target
2. Determination of target structure
3. Identification of binding site
4. Computational drug design methods
5. Evaluation of potential lead candidate

2.2.1 Identification of Drug Target

A drug target or a receptor is a macromolecule that is critical for a disease condition to happen. The identification of a drug target starts with understanding the structural biology of the target. Function prediction, pathway information, disease associations, and structural data need to be considered when choosing a drug target. Different bioinformatics tools are used to gather this information, helping to choose the potential target that is critical for the disease function.

The objective in drug development against pathogenic diseases is total inhibition leading to the death of the pathogen. The target should be located in a critical step of a metabolic pathway and its inhibition should consequently kill the pathogen. Drug targets should be unique so that no other pathway should suppress the impact of the inhibitor. One important criterion of a drug is specificity. The physiological effect of the drug should be as clearly defined as possible. It should specifically bind to the target protein in order to minimize undesired side-effects.

2.2.2 Determination of Target Structure

Accurate structural information is needed after the drug target is chosen. Knowing a protein's structure is vital to understanding how it functions. Following are three methods by which structural data is obtained.

1. X-ray crystallography (XRC)
2. Nuclear Magnetic Spectroscopy (NMR)
3. Homology Modeling or Comparative Modeling

X-ray crystallography (XRC) is an experimental technique in which X-rays are diffracted by crystals. XRC is the main source of information for drug design as it gives high resolution structures. Once the 3-dimensional structure of protein is determined, the co-ordinates of the atoms are used by the computer to construct the protein structure.

Nuclear magnetic resonance (NMR) is a spectroscopic technique that uses magnetic fields and electromagnetic radiation to determine the structure of organic compounds. NMR is commonly used for characterizing the structure and molecular dynamics of target or ligand molecules.

Homology modeling finds the approximate 3-dimensional structure of the target (not determined either by X-ray or NMR technique) based on an available sequence, provided with an empirical structure template with at least 30% sequence identity. This method can be applied to generate reasonable 3-dimensional models of protein structures.

2.2.3 Identification of Binding Site

After structure of the target is determined, the binding site (also known as the active site) needs to be identified. The ligand is a small molecule which can be ingested orally. The binding site is a hollow cavity-like structure and has hydrogen bond donors as well as hydrophobic characteristics. When the ligand binds to the target at the binding site and is capable of inhibiting the target, then the ligand is considered a lead compound or lead candidate.

2.2.4 Computational Drug Design Methods

The goal of all computational studies used in drug design is to evaluate the inhibitor binding complex. After identifying the possible binding regions on a target structure, potential ligands that bind and inhibit the target are evaluated. Computational and experimental methods are followed for finding a good lead. Computational methods for structure-based drug design are not mature enough to replace existing practices, but new approaches and computer programs continue to appear. Most work falls into the categories of *De novo* design, Docking, and Scoring.

De novo design involves designing of novel ligands intended to bind to a given receptor with high affinity. Researchers split the problem of ligand binding into two

sections; docking and scoring. For a known target structure and a known ligand, docking involves locating the bound configuration of the ligand and receptor, while scoring involves evaluating the affinity of the ligand, given a particular bound configuration. Docking and scoring are complementary methods in structure-based drug design.

2.2.5 Evaluation of Potential Lead Candidate

Once the ligand has been successfully identified, it is called the lead compound. It must be evaluated for binding affinity before proceeding to the preclinical trials. The lead is evaluated for bioavailability, oral viability, chemical and physical stability, and ease of production.

CHAPTER THREE

LITERATURE REVIEW

The field of drug design is a widely researched area with many successes since 1980's. Every step in drug design requires tremendous research effort. This literature review discusses the popular methods and programs that were used to discover drug targets and to determine target structure. Computational drug design methods and concepts along with advantages and disadvantages are presented in this review for detecting ligand binding sites and docking the targets. The merits and demerits are also discussed in this section.

3.1 Identification of Drug Target

The drug targets in structure-based drug design are mostly proteins. The proteins belong to families of kinases, proteases, and peptides. Enzymes are exceptional drug targets because compounds can be found to fit into the active site.

The most popular method for identifying drug targets is by the study of the disease under observation and by identification of genes associated with the disease. Coombs et al. in their review state that the *P. falciparum* Genome Project had identified several drug targets for malaria. They researched hemoglobin metabolism and found that enzymes called plasmepsins of *Plasmodium falciparum* (mosquito) were involved in hemoglobin degradation. They found that Plasmepsins 1, 2, 4 and Histo Aspartic Protease (HAP) played a significant role in hemoglobin degradation and were potential drug targets [5]. Banerjee et al. suggested that proplasmepsin convertase could also be used as a promising drug target because it produces plasmepsins [6].

Genomics related methodologies are applied for choosing drug targets. Sakharkar et al. used three-way genome comparisons to identify essential genes from *Pseudomonas aeruginosa*. The National Center for Biotechnology Information (NCBI) and Database of Essential Genes (DEG) databases were used for downloading the protein sequences for *P. aeruginosa* and humans. The *P.aeruginosa* genes were eliminated at 60% using pathogen genes without homologs in humans. An expectation (E-value) cutoff of 10^{-3} was used and the non-homologous entries were then subjected to BLASTP (Basic Local Alignment Search Tool) against the DEG database for the identification of homologs to essential genes at a cutoff score of 10^{-10}. Their approach identified 306 essential genes that may be considered as potential drug targets for new antibiotics [7].

3.2 Structure Determination

The 3-dimensional structure of HIV-I protease was extensively studied and characterized through XRC studies. It was the first breakthrough that made rational drug design for AIDS possible. The HIV protease structure was first determined in 1987 by Pearl et al. through homology modeling using the known structures of eukaryotic aspartic proteases as templates. They used structure prediction and molecular modeling techniques, concluding that the viral protease sequences correspond to a single domain of an aspartic protease, and that this domain may function in a dimeric form. The researchers also constructed a model of the pol-protease of HIV-1 to test this hypothesis [8].

Navia et al. from Merck laboratories achieved HIV protease structure in 1989 [9]. A more accurate model was reported by Kent and his co-workers from the National

Cancer Institute. They synthesized an active protease entirely out of D-amino acids that was able to catalyze the proteolysis of D-peptides [9]. Figure 1 shows the structure of HIV protease.

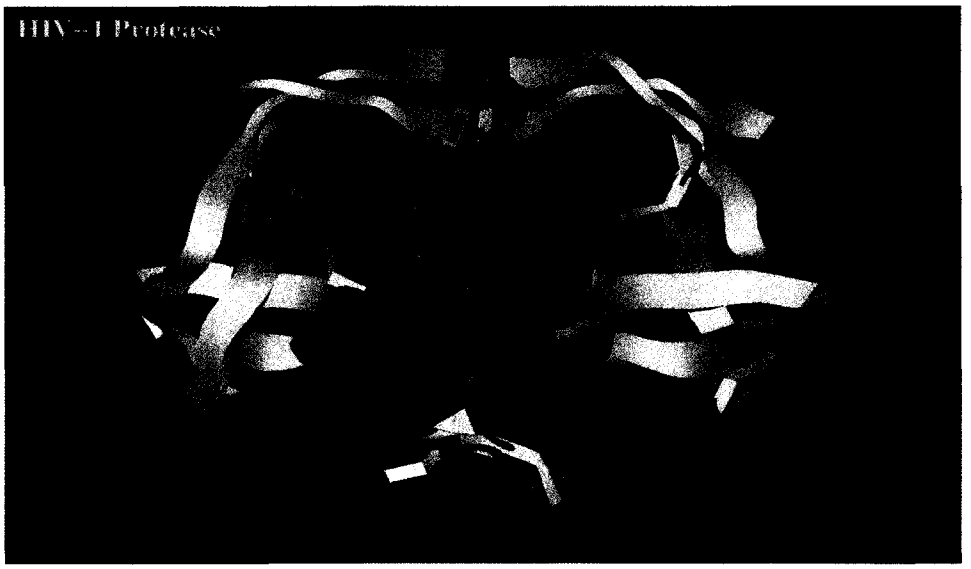


Figure 1. Structure of HIV protease [Reprinted with permission from Dr.T.N.Bhat, "HIVdatabase PR gallery", available at http://xpdb.nist.gov/hivsdb/gallery.html].

In its mature form, the viral protease exists as a dimer, whose subunits each consist of 99 amino acids. The folded subunits together interact to form a hydrophobic core and two flexible flaps that can close around the substrate [10].

The above research work shows clearly that XRC and Homology Modeling techniques played a major role in structure determination. Currently there are more than 140 HIV protease structures available in the online database of the National Cancer Institute. The RCSB protein data bank currently has 40,354 biological macromolecular structures stored in the database [11]. Most researchers download structures for their study from this information portal for molecular modeling purposes.

3.3 Identification of Binding Site

Kuniyasu identified the binding sites using the photoaffinity labeling method. The binding sites of four typical calcium antagonists, 1, 4-dihydropyridines, benzothiazepines, phenylalkylamines, and benzothiazines, were successfully identified within the primary structure of the skeletal muscle calcium channels [12].

Co-crystallization studies in which the target molecule is crystallized with a small bound molecule inhibitor can be valuable to determination of the target site. Filikov et al. identified three ligand binding sites for RNA targets in HIV-1 TAR using virtual screening [13].

Computational tools have been developed to detect the ligand binding sites or the depressions in the target surface in order to localize the potential binding sites. These methods rely on purely geometric criteria with different algorithmic approaches. There are three cavity identification programs, namely: LIGSITE [14], APROPOS (Automatic PROtein POcket Search) [15], and PASS (Putative Active site with Spheres) [16].

Hendlich et al. developed LIGSITE which embeds the protein in a uniformly spaced grid. Lattice intersections that coincide with the protein atom's Van der Waals sphere are discarded. The remaining lattice points are scored according to their degree of burial in surface depressions [14, 17].

Though all the above cavity identification programs are freely accessible to all, most of the docking software provides similar programs as an accessory. In addition, most of the binding studies are available in scientific journals and databases. With the

information provided in the databases, there is less need to use the cavity identification software.

3.4 Computational Drug Design

De novo design is the generation of new compounds that do not exist in compound databases. This means that the ligand structure is not known. Many programs were created for de *novo* drug design. Bohm developed a new method for *de novo* design of inhibitors called LUDI which can attach a new substituent into an already existing ligand. The application was used for binding HIV protease inhibitors and inhibitors of dihydrofolate reductase. The LUDI algorithm suggested modifications to increase the binding affinity between an existing ligand and target [18, 19].

The LUDI algorithm represents a molecular-fragment based approach to identifying potential *de novo* leads. Fragments are identified that make hydrogen bonds with the target and fill hydrophobic pockets on the target, and the fragments are then connected with linkers to make a single molecule. The matching process is performed using the fragment library for the binding site. The matching fragment is then docked for alignment. The method was tested with dihydrofolate reductase (DHFR) and trypsin, and was found to generate fragments that aligned favorably with known inhibitors such as methotrexate and benzamidine [18, 19].

The other *de novo* design programs are SPROUT [20], HOOK [21], Growmol [22], GRID [23], CAVEAT [24], and Legend [25]. These programs generate new structures that fit the site of the target, but have two common critical defects; the synthetic feasibility of the novel structures is not considered and the prediction of binding

affinity is not accurate. These defects make it difficult to obtain the new compounds from *de novo* generation. One of the defects (prediction of binding affinity) can be overcome by using the scoring function. The formula is as follows:

$$- RT \ln K_d = \Delta G^\circ = \Delta H^\circ - T \Delta S^\circ \qquad \text{Equation 1}$$

Where R is the gas constant at the absolute temperature T, K_d is the dissociation constant of a ligand-target complex and measures binding affinity, and ΔH° and $\Delta S^\circ$are the changes in standard enthalpy and entropy. ΔG° is the change in standard Gibbs free energy.

The principle of free energy perturbation and molecular dynamics simulation was reported by Kollman et al. Free energy perturbation (FEP that could be used to calculate binding affinity for similar compounds) and molecular dynamic simulations were both time consuming [26]. The researchers, in order to overcome this defect, developed the generalized Born model and the Ligand Interaction Energy (LIE) for computational speed and accuracy of binding affinity.

The interaction of inhibitors with HIV protease was the beginning for various docking methods. The first docking method was developed by Kuntz et al. They discovered that haloperidol could be used as an inhibitor for the HIV protease. As the structure prediction of the inhibitor turned out to be different from the predicted one, haloperidol was not used [27, 28].

To predict binding sites in a target, DOCK software was used. DOCK software predicted the binding modes of small molecule-protein complexes. First, the potential sites of interest on the target molecule are identified. The molecular surface area of the

target is generated and a negative image of the surface is created around the active site. Sphere centers are then matched to the ligand atoms, for possible orientations. Typically, tens of thousands of orientations are generated for each ligand molecule. The orientation of the ligand is evaluated using scoring grids. Scoring grids are pre-calculated energy information for each atom of the target. The ligand-receptor binding energy is the sum of van der Waals attractive, van der Waals dispersive, and columbic electrostatic energies. As a result, interpolated receptor values are obtained. Finally when docking is performed, the ligand values are substituted and the orientation of the ligand is evaluated based on the lowest energy score for effective binding [29].

DOCK runtimes depend on many variables, including the number of ligands being explored, ligand size, target size, and sampling parameters. It can take anywhere from a few seconds for a single ligand to a number of days for large databases. The following page shows how DOCK works with Figure 2, Figure 3, and Figure 4.

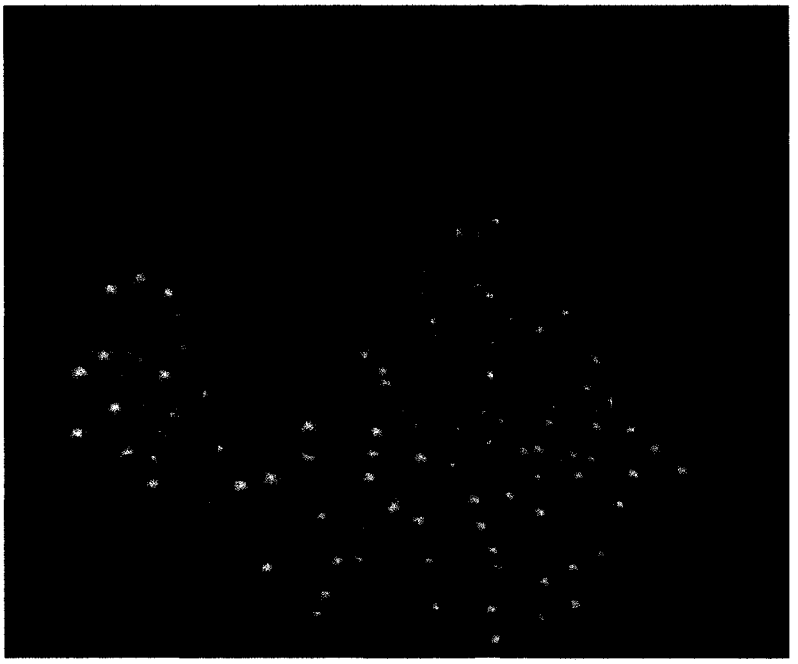


Figure 2. Spheres generated over negative image of receptor surface.

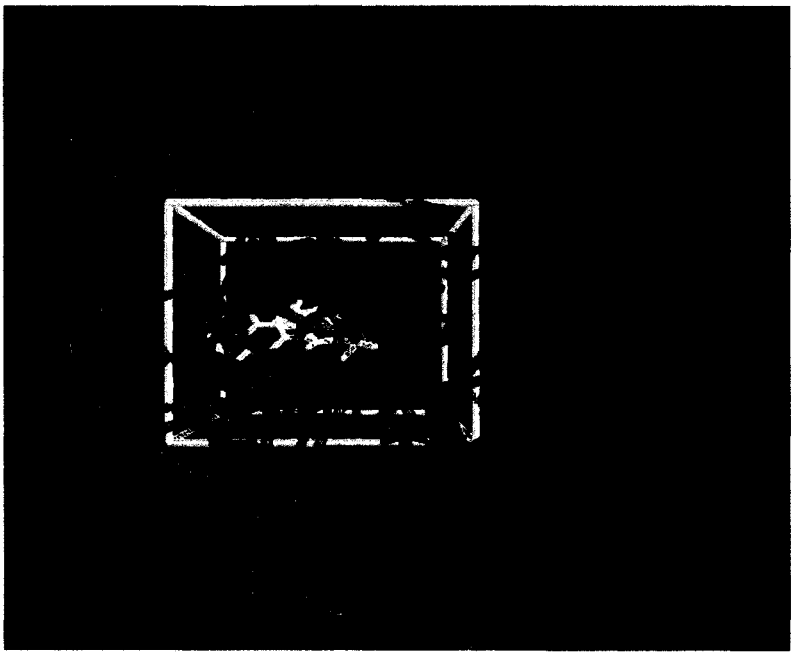


Figure 3. Energy grid for target is calculated.

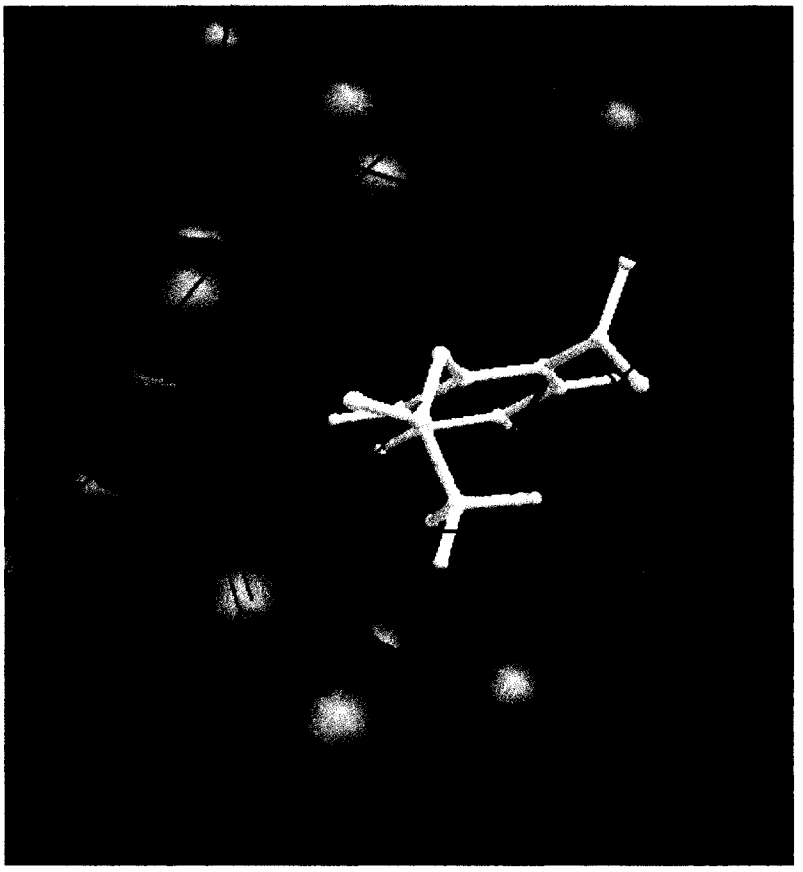


Figure 4. Docked molecule with multiple ligands.

3.5 Drug Lead Evaluation

Lipinski's Rule of 5 is followed for the lead drug evaluation [30]. It is called Rule of 5 because all the four parameters evaluated are close to five or a multiple of 5. Lipinski states that poor absorption or permeation is more likely when

- There are more than five H-bond donors.
- The molecular weight is over 500 Da.
- The logP (partition coefficient) is over 5.
- The sums of N's and O's are over 10.

Veber et al. state that the number of rotatable bonds should be less than 10 and the H-bond donors or acceptors should be less than 12 for oral bioavailability [31].

3.6 Summary

Structure-based drug design is a powerful method for discovering new drugs against diseases. From the literature review we see that XRC and homology modeling are excellent tools for structure determination. The structures stored in the databases are an ideal choice for selecting and downloading the structures of the macromolecule. The same applies to the ligand as well. In the case of target site identification and drug target, literature search is the best option; there is a wealth of information available on the internet and the electronic databases. Purely geometric tools for binding site identification are generally well suited to localize all significant cavities or depressions on the protein surface. The UCSF DOCK software is found to be a simple, effective and representative tool. For the ligand evaluation stage, the Lipinski's Rule of Five and the number of rotatable bonds are used.

CHAPTER FOUR

RESEARCH OBJECTIVES AND JUSTIFICATIONS

4.1 Research Objectives

The research objective of this thesis is to develop a laboratory module in order to introduce basic principles of drug design. The module is intended to be used in a graduate level, bioinformatics or bioengineering class as lab work in the chemical engineering department of SJSU. It is important for students who aim to enter into a biotech industry to have basic practical knowledge of drug design.

The bioinformatics lab course covers practical and theoretical concepts of genomics, proteomics, and relevant bioinformatics tools. Students will work on the drug design lab module based on techniques and skills they have learned throughout the semester. This exercise will help them get familiar with the bioinformatics tools employed in various applications. To understand drug research concepts, there will be five main modules based on the process of structure-based drug design. The key goals of the module will be

- To identify the drug target
- To visualize the drug target
- To find ligand candidates that uniquely interact with the target
- To dock the ligand with the drug target
- To evaluate the docked ligand

4.2 Justifications

- The faculty of chemical engineering at SJSU suggested incorporating a drug design module into the bioinformatics course as students only learned basic practical concepts to a limited extent without it.

- The module has been carefully developed. It is cost effective, as the chosen software is free to academic users.

- The students do not require any programming skills, except the knowledge of UNIX commands, as the drug design module is very user-friendly.

Even though the drug design module is brought together in a simple mode, it is still an ambitious effort. This module will evoke interest among students once they complete it; hence, they will further develop the module or use the tools in the module to facilitate their own bioinformatics research.

CHAPTER FIVE

METHODS AND MATERIALS

A student module was developed to give students an exposure to the drug design process. The module includes steps to follow for the procedure and the activities. The activities contain a small research project and questions at the end of the module to help students understand the process of drug design. The drug design lab module is presented in Appendix B. This section explains the materials and methods required to develop the module.

5.1 Module Flowchart

Figure 5 gives an overview of the structure of the modules.

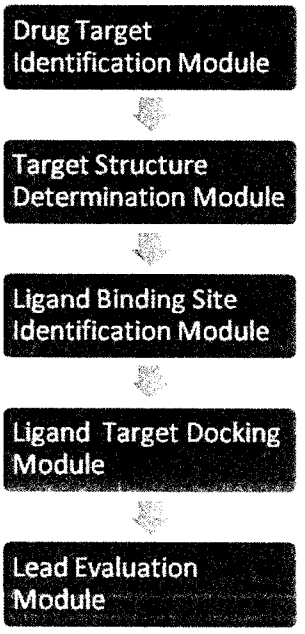


Figure 5. The drug design module flowchart.

5.2 Materials Required

All the materials and software for the modules are readily available. The following information was required to develop and work the module:

1. An uncomplicated human disease and its pathway
2. The drug targets associated with the human disease
3. A 3D molecular graphics program-VEGA ZZ
4. The literature information regarding binding sites
5. A basic docking software- UCSF DOCK
6. The lead evaluation criterion

The constant factors used in this module are the disease and the chosen drug target. The variable factors are the ligands. The names of the ligands are not disclosed to the students so that they can evaluate them purely based on their structures. A 3D molecular graphics software called VEGA ZZ and docking software called DOCK were used. The complete information for the software can be obtained on the internet.

As for the module, a clear set of instructions has been provided along with the background information. A set of questions and activities has been developed at the end of each module. The questions and activities aim to make the students understand the process of drug design and encourage them to use bioinformatics tools.

5.3 Drug Target Identification Module

In this module the students' objective is to evaluate the drug targets in relationship to the disease and identify the best target. Information regarding the drug targets, such as function, disease pathway, and background, is presented to them. The

disease pathway provided to them will guide them indirectly towards the right answer. This module helps them understand the process of identifying a proper target for a specific disease and allows them to get a foothold in the design process.

5.4 Target Structure Determination Module

The student objective of this module is to determine the 3-dimensional structures of the target. This involves structure visualization and validation. A simple tutorial was developed for students to determine the target structure with the help of a molecular graphics program. Determining the target structure and viewing it will help the students understand the components of the target on a molecular level. Questions at the end of the module will help students understand the structural details.

5.5 Ligand Binding Site Identification Module

The student objective of this part of the module is to find potential ligand binding sites in the chosen drug target structure. Since there was considerable information about the active site presented in scientific journals, no software was required. The students are asked to do a literature search to find the active site(s) and their exact location using the electronic research databases provided in the SJSU library. This module will help them analyze the active site of the target structure in great detail.

5.6 Ligand and Target Docking Module

The student objective was to bind the ligand with the target structure geometrically using the docking software. The ligands were downloaded from the KEGG (Kyoto Encyclopedia of Genes and Genomes) Ligand Database website. This

module will help the students learn and appreciate the complex of two molecules via docking. This module is comprised of four stages outlined below.

- Molecule preparation for ligand and receptor
- Sphere generation of the receptor
- Scoring grids for the receptor
- Docking the receptor and ligand

The manual provided by docking software was followed for this module. Students will perform ligand protein docking and will evaluate the parameters needed for this module.

5.7 Potential Lead Evaluation Module

The student objective of this module is to evaluate the ligand and to find the best lead. An evaluation criterion to analyze ligands is listed in the module. It is up to the students' discretion to find a lead and discuss why it is a good one.

It is not unusual to obtain erratic results and conclusions due to many factors. The initial target selected may not have been a good one; in this case another target may be selected. The target site identification and analysis may not be correct. This is easily rectified by picking another target site. If and when this happens, the students will re-evaluate their drug target or find another ligand binding site and will perform another set of experimental runs.

CHAPTER SIX

RESULTS AND DISCUSSION

The disease state taken up for study in this module is Osteoarthritis. This disease affects almost 21 million adults, in the United States alone. Osteoarthritis was chosen because it has several drug targets with specific function and pathway information that is simple to understand. Many drugs have been found that are used to treat the disease, and researchers are still in the process of finding a cure. Numerous targets are associated with arthritis ranging from enzymes to hormones, chemokines, and cytokines. All these target molecules generate pain signals. A lot of structural and functional information about target molecules are essential for choosing the right target upon which the drug can act.

6.1 Background of Disease in Study

Osteoarthritis is a chronic disease of the joints involving breakdown of the joint tissue, primarily the cartilage. Cartilage is a tough elastic tissue covering the bone that acts as a shock absorber and keeps the bones from rubbing against each other. The entire joint is enclosed in a capsule that is lined by an inner membrane known as the synovial membrane. The membrane fills the space around the joints with a fluid called the synovial fluid. The synovial fluid nourishes the cartilage and keeps the joints lubricated making movement smooth and easy. Enclosing the joints are muscles, tendons and ligaments. These structures provide support and assist movement in the correct direction. Figure 6 shows the difference between a normal joint and a degenerative joint.

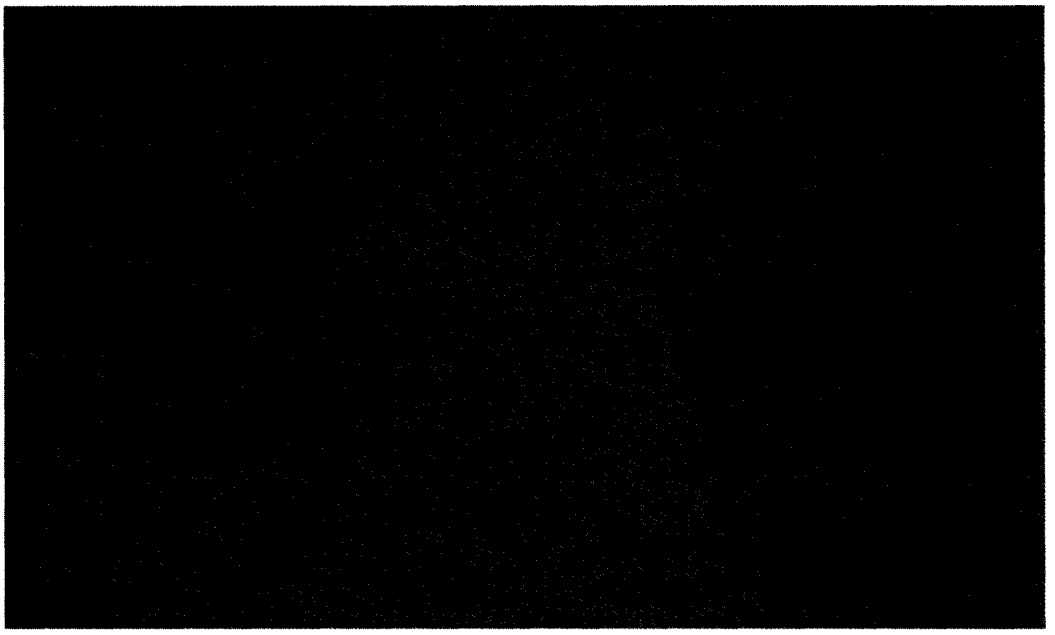


Figure 6. Comparison of a normal joint and osteoarthris joint[diagram by author].

The degeneration of the cartilage causes inflammation. Inflammation is one of the ways in which the body protects and repairs itself. The inflammatory reaction begins with tissue irritation. In response to tissue irritation, white blood cells rush from the bloodstream to the area in order to facilitate the healing. They release enzymes and active products that affect the nearby cells and alter the blood circulation. The active products are prostaglandins and leukotrines. They are the potent producers of inflammation. They increase the flow of fluids and white blood cells around the area of inflammation and release more prostaglandins thereby feeding the process. The accumulation of these fluids and cells causes swelling along with pain [32].

There has been a lot of research regarding the cure of osteoarthritis. To date, there has been no clinically approved drug to cure it. Pain relievers such as analgesics

and NSAIDs (Non-Steroidal-Anti-Inflammatory-Drugs) are used for temporary pain relief. These medications cause as many side effects as benefits. Treatment for osteoarthritis is aimed at controlling pain, improving and maintaining movement, and preventing degeneration of the joints. Pain control is a vital part of the treatment as the drugs help the patient gain control over the effects of the disease. Rest and exercise help in maintaining movement and preventing degeneration of the joints.

The drug targets that were chosen for study (along with the pathway description) are as follows.

1. Cyclooxygenase enzyme-1 (COX-1)
2. Cyclooxygenase enzyme-2 (COX-2)
3. Prostaglandins
4. Interleukin-1 receptor

Prostaglandins are hormones that act as chemical messengers inside the cell activating inflammatory responses and strengthening pain signals. They are unsaturated carboxylic acids consisting of a twenty-carbon skeleton and a five-membered ring. They are synthesized from arachidonic fatty acid. In physiological conditions archidonic acid exists in the ionic form, hence is named arachidonate. This initial reaction in their generation is catalyzed by enzymes called cyclooxygenases (COX-1 and COX-2). The prostaglandins are essential for cytoprotection in the stomach, platelet aggregation and renal function. Prostaglandins produced by COX-2 cause inflammation.

There are two different cyclooxygenases (COX-1 and COX-2). Both the enzymes perform the first step in production of prostaglandins by adding two oxygen molecules to

arachidonate. COX-1 and COX-2 enzymes shown in Figure 7 are similar in structure and have the same substrate (arachidonate). The kinetics of both the isoforms are similar but they obtain arachidonate from different sources.

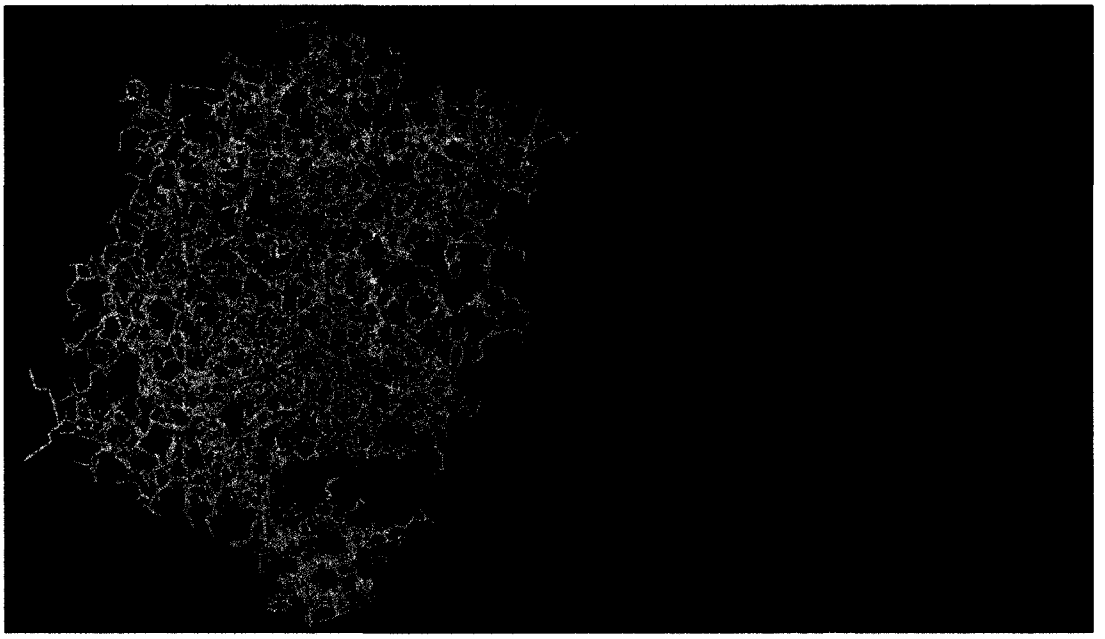


Figure 7. The structure of COX-1(cyan) and COX-2(magenta) enzymes.

The COX-1 enzyme is a constitutive enzyme, is present in all cells in the endoplasmic reticulum and is maintained at a constant level. The COX-1 enzyme is a "housekeeping enzyme" involved in signaling, tissue homeostasis and cytoprotection. Most importantly, it produces prostaglandins in the stomach to help maintain mucosal epithelium, and its inhibition cause gastric damage, bleeding, and ulcers. COX-1 initiated prostaglandins are required to maintain normal renal blood flow in impaired kidneys. Prostaglandins synthesized with the aid of COX-1 are essential for platelet function.

The COX-2 enzyme is an induced enzyme. An inducer can increase its concentration in the range of 10–80, however this is reduced considerably after a few hours. It is found in high concentrations around macrophages, synovial cells, mucocytes, and fibroblasts in response to inflammation. The COX-2 enzyme is characteristic of increased inflammation when the system is trying to heal the inflamed area.

The fourth drug target is the Interleukin-1 receptor that belongs to a family of cytokines. Cytokines are small secreted proteins which are pro-inflammatory. They contribute to the bone degeneration characterizing arthritis. The Interleukin-1 receptor is present in the synovial lining of the joint. The main function of this particular cytokine is to cause inflammation by activating monocytes and macrophages. Interleukin-1 stimulates the production of prostaglandins and nitric oxide, and promotes joint degradation. The expression of Interleukin-1 can be stimulated by various types of cell interactions such as TNF, autocrine or paracrine. When one type of cytokine is inhibited the other cytokines replace it and carry on its function. Other cytokines having pro-inflammatory properties or catabolic factors could also contribute to this pathological condition, and those that have anti-inflammatory properties may be able to counteract the negative effects of the former on the disease process.

6. 2 Choosing the Right Drug Target

There are several possible modes of drug-based therapy. These include targeting prostaglandins, the two enzymes COX-1 and COX-2 associated with prostaglandins synthesis, and the Inter-leukin receptor. The module helps students choose the right drug target based on available literature information. The analysis is as follows:

When analyzing the given drug targets, it is apparent that prostaglandins which are responsible for pain cannot be the target because their production is initiated by two different cyclooxygenases which generate different classes of prostaglandins. The prostaglandins produced by COX-1 are essential for cytoprotection in the stomach, platelet aggregation and renal function. Prostaglandins produced by COX-2 cause inflammation. If prostaglandins are inhibited it would not only stop inflammation but also stop positive functions of the body and cause damaging side effects. Hence prostaglandins cannot be used as a drug target.

Cytokines cannot be used as a drug target because specific blockage of the Interleukin-1 receptor causes other cytokines to replace its function. Therefore, it would not be a worthwhile strategy to inhibit the Interleukin-1 receptor.

It is evident that the COX-1 enzyme also cannot be used as a drug target because inhibiting this cyclooxygenase isoform causes gastric damage, hemorrhaging, and ulceration in the stomach. The inhibition also causes irregular blood flow in the kidney and platelet dysfunction in the circulatory system. COX-1 (present in all cells) is important for cell functions such as homeostasis.

Although COX-1 and COX-2 enzymes are 60% similar in structure they carry out different functions; one maintains normal, cellular physiology while the other initiates the generation of pain. The COX-2 enzyme is up regulated during inflammation and is usually present in insignificant amounts in the human system; hence its inhibition would not affect the regular functions of the body. Since the aim of treatment is pain control

with few side effects, the COX-2 enzyme is the ideal drug target. Inhibition of the COX-2 enzyme would result in anti-inflammatory and analgesic effects.

6. 3 Determination of Target Structure

After the students choose the right target, the next step is to determine the target structure. The students are required to visualize and obtain first hand information about the target in study. The drug target for this case study is cyclooxygenase-2 (COX-2). The structure of COX-2 is visualized using VEGA ZZ, an interactive molecular graphics program. VEGA ZZ is a new molecular modeling program widely used and free of charge for academic users. It is an interactive, user friendly program mainly used to visualize biomolecules. Additionally, it performs tasks like adding hydrogen, deleting solvent, adding charges and changing the file format. Since the modeling program is also useful for docking purposes, it has been used in the fourth stage of the module as well. The software is available online and can be downloaded and installed from the website http://www.ddl.unimi.it. UCSF Chimera is another similar molecular modeling program that has been used scarcely.

The COX-2 crystallographic structure is obtained from RSCB protein data bank which contains information about macromolecular structures in pdb format. Every molecule is denoted by a pdb id. An uninhibited mouse COX-2 structure is used as the target. The corresponding pdb id is 5COX.pdb. This target structure (5COX) is used because it is similar to that of the human COX-2 structure (87% similarity) and a compatible human COX-2 structure was not available in the database [33]. Most

researchers have worked with the mouse structure in the drug design phase. The structure of cyclooxygenase-2 is viewed in the VEGA ZZ program shown in Figure 8.

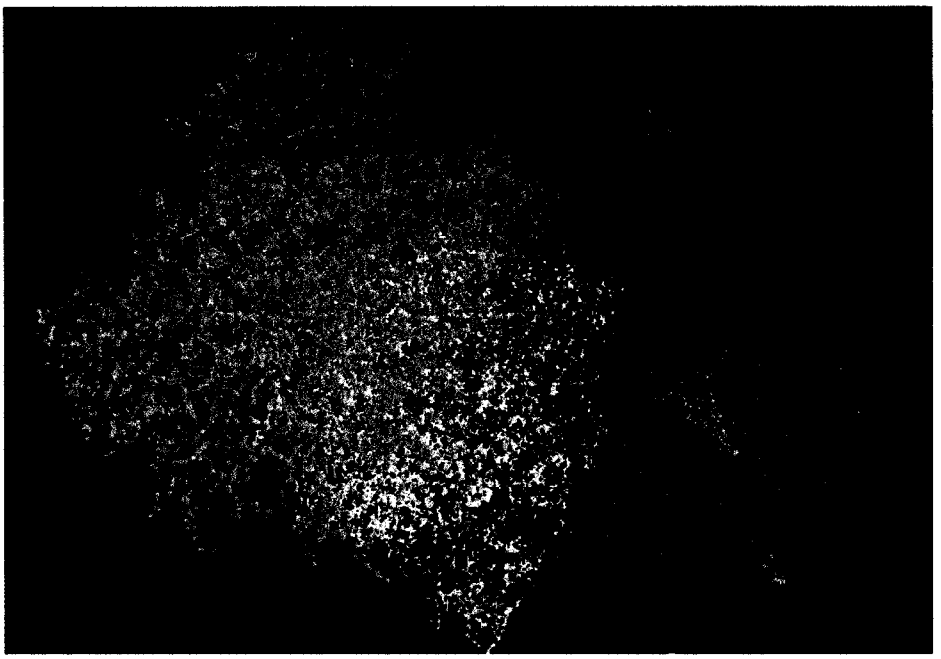


Figure 8. The structure of COX-2 enzyme.

The basic characteristics of COX-2 are listed below.

- The COX-2 structure is a homodimer of identical subunits. It is comprised of four protein chains and 587 amino acids. The four protein chains are indicated in blue, green, red and white shown in Figure 12. The structure contains four heme ligands and 12 N-acetyl glucosamine (NAG) trisaccharide ligands shown in yellow in Figure 8.

- The heme group is protoporphyrin IX containing iron (Fe). The metal free porphyrin combines with the ferrous ion to form the heme group.The enzyme has a cyclooxygenase active site and the peroxidase active site separated by a heme prosthetic group.

6.4 Determination of the Target Active Site

After the determination of the target structure, the students are led to the third stage of determining the active site of the target. There is ample information about the

location of the active site and its function on the internet. The structure of the drug target is loaded into VEGA ZZ and with the information given from the journals the active site is located. Since there are two active sites, students are indirectly guided to find the correct active site. A literature search was performed in the electronic database provided by the SJSU library. The keywords used were "COX-2 + active site + location." Figure 9 shows the location of the active site enclosed by residues.

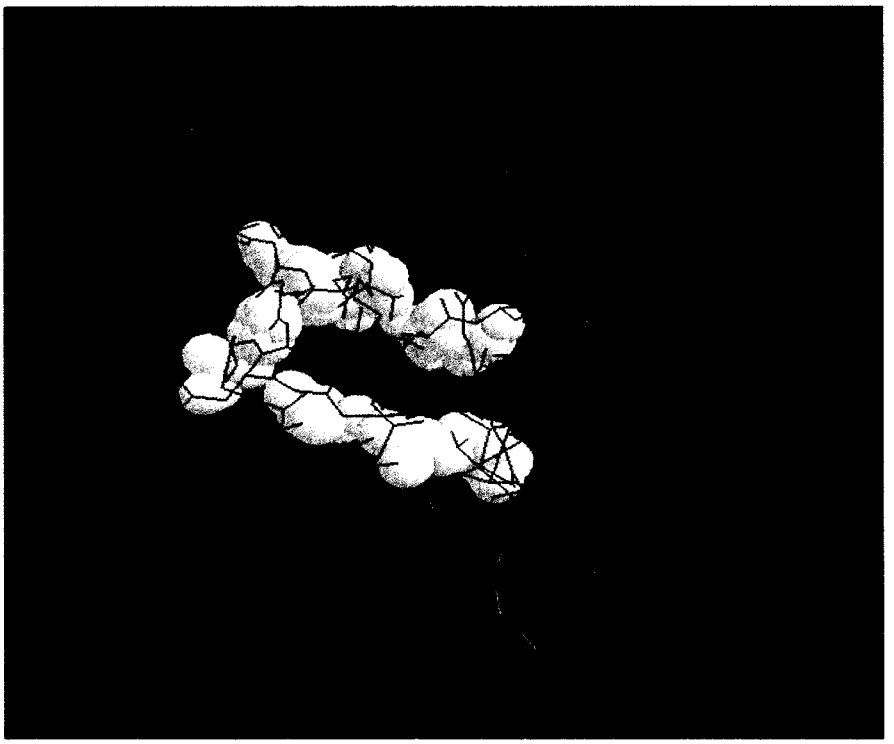


Figure 9. The COX-2 structure with active site surrounded with white residues.

The COX-2 enzyme structure has two different active sites, the cyclooxygenase active site and the peroxidase active site. At the cyclooxygenase active site, arachidonate binds to form prostaglandin (cyclopentane hydroperoxy endoperoxide; PGG2). If

arachidonate is prevented from binding, then prostaglandins will not be formed, hence no inflammation will occur through COX-2.

The peroxidase active site converts the peroxide in PGG2 into an OH group transforming PGG2 to PGH2. This active site is needed to produce a tyrosyl radical in newly synthesized COX (Tyr 384 in COX-1 and Tyr 371 in COX-2) that participates in the cyclooxygenase reaction. This tyrosyl radical is radical is critical to the reaction catalyzed by cycloxygenase and is regenerated by the catalytic well. The cyclooxygenase active site is chosen to be the right active site and will be used for the docking stage.

The cyclooxygenase active site is defined by three different regions. There is a hydrophobic pocket area surrounded by residues tyrosine385, tryptophan 387, phenylalanine 518, tyrosine 248, and leucine 352. The entrance of the active site is defined by arginine 120, Glu 524, and tyrosine 355 and there is a side pocket enclosed by valine 523, arginine 513 [34]. Since the valine molecule is small, it leaves a gap causing a side pocket. This side pocket is the site of binding for many selective drugs.

6.5 DOCK

After analyzing the COX-2 active site, the students are guided to the fourth stage called the docking stage. In this stage molecular docking software called DOCK is used. The students will be given reference instructional manual to operate of the software and will proceed with the docking process.

The UCSF DOCK software was chosen for docking purposes. The manual provided by DOCK was followed for the docking procedure. Test runs for ligand protein docking were performed and the parameters needed for this module were evaluated.

6.5.1 Overview

The DOCK software predicts the binding modes of small molecule-protein complexes. The docking part is divided into series of steps. First, the potential sites of interest on the target molecule are identified using the negative image of the target structure. The sites of interest include the active site as well as the other potential binding sites. This means it does not differentiate between the actual active site and the other potential sites.

Secondly, the software generates spheres which fill the sites of interest chosen by DOCK. The sphere centers are assumed to be ligand atom positions. The spheres touch the surface of the molecule but do not intersect it. The spheres are allowed to overlap or intersect other spheres. Figure 10 shows how spheres are arranged in a sample binding site. The blue spheres are generated with the help of points from molecular surface indicated by the green line [35].

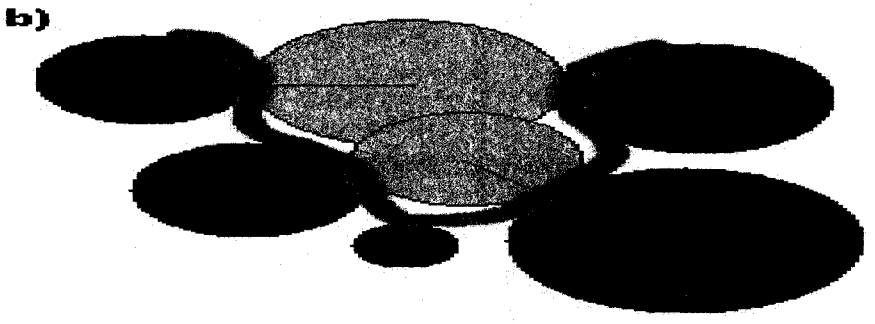


Figure 10. Representation of the spheres arranged in a binding site. [Reprinted with permission from Dr.Irwin Kuntz and Dr.Therese Lang, "Tutorial: generating spheres", http://dock.compbio.ucsf.edu/DOCK_6/tutorials/sphere_generation/generating_spheres.htm]

After the spheres are generated, the sphere centers are "matched" with the ligand atoms in order to position a ligand in the active site. Many such sphere-ligand atom sets are formed using the longest distance heuristic method. The entire orientation of the ligand atom within the active site is calculated using sphere-ligand atom sets [36].

Thirdly the ligand orientation is evaluated using scoring grids. Scoring grids are pre-calculated energy information for each atom of the target. Every atom of the target has an associated ligand-target binding energy and an electrostatic charge. As a result, interpolated receptor values are obtained. Finally when docking is performed, the ligand values are substituted and the orientation of the ligand is evaluated based on the lowest energy score for effective binding [36].

There are two types of docking, flexible and rigid. In rigid docking, the receptor is assumed to be a rigid structure that does not flex itself to fit the ligand [36]. In flexible docking, it is assumed that the receptor structure can change its shape to align with the ligand during docking.

A standard method for docking described in the UCSF DOCK manual was followed. The starting data are COX-2 and ligand structures. To dock the target structure a step by step procedure was followed.

1. Molecule preparation for target and ligand
2. Sphere generation of the target
3. Scoring grids for the target
4. Docking the target and ligand

6.5.2 Molecule Preparation for Target and Ligand

The purpose of this step is to prepare the target and the ligand molecule as an input to DOCK. For the target structure (which is a homodimer) only one subunit of COX-2 is required. Using the "Remove Segment" option in VEGA ZZ, all chains except one are deleted. Figure 11 shows a single chain structure of COX-2.

Figure 11. Single chain of COX-2 structure.

With the help of the "edit" option in VEGA, minor modifications are executed. The minor changes are solvent deletion, hydrogen addition, and charge calculation. Solvent deletion is used so that DOCK software can explore the structure volume. Solvents such as water, as well as cofactor and ions are removed. Hydrogen atoms are added to the receptor structure to study different ligand interactions. The file is saved as

5coxrec_charged.mol2 and is shown in Figure 12. Hydrogen atoms are deleted from the file for molecular surface generation and the file is saved as 5coxrec_noH.pdb.

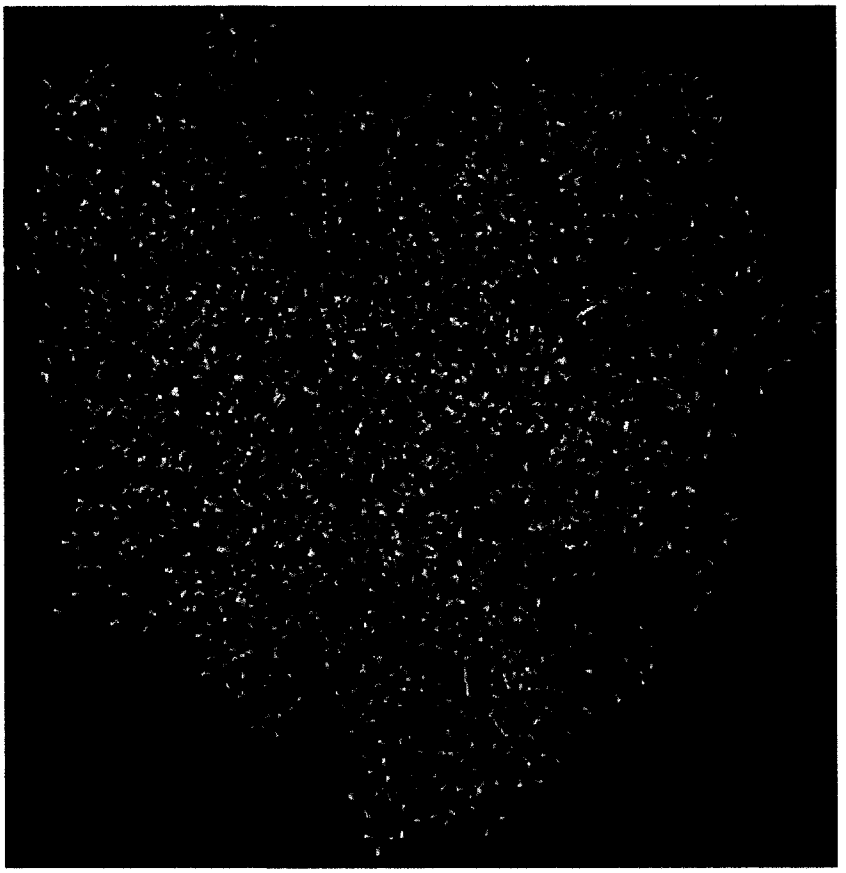


Figure 12. The structure of 5coxrec_charged.

6.5.3 Generating Spheres for the Target

As a prerequisite for generating spheres in the target structure, the molecular surface of the target is generated using the dot molecular surface (dms) program. The input for the dms program is 5coxrec_noH.pdb. The output file is saved as 5coxcutto50_more.ms and is shown in Figure 13 in the pdb format.

Figure 13. Molecular surface of the target (5coxcutto50_more.ms).

To generate spheres over the molecular surface of the target, a program called sphgen is used. A file called INSPH is generated in the format shown in File 1. The file contains 5coxcutto50_more.ms, parameters, and an output file name (5coxrec.sph).

File 1. The input file INSPH for sphere generation.

```
5coxcutto50_more.ms
R                              (sphere outside of surface)
X                              (subset of surface points)
0.0                        (prevents generation of large spheres
                                with close surface contacts)
4.0                        (maximum sphere radius in Angstroms)
1.4                        (minimum sphere radius in Angstroms)
5coxrec.sph                                    (output file)
```

The command "sphgen" is run in the same folder that contains the INSPH file. Two output files are generated of which one is the 5coxrec.sph, which contains spheres in clusters. The output file generated has 49 clusters for the target molecule. The second output file is called the OUTSPH and contains calculation information. The 5coxrec.sph file is converted into pdb file format for viewing. File 2 shows the output file OUTSPH, and File 3 shows a partial file of 5coxrec.sph.

File 2. The OUTSPH file.

```
density type = X
 reading  5coxcutto50_more.ms
type    R
 # of atoms =    2635
 # of surf pts =  92378
 finding spheres for   5coxcutto50_more.ms
 dotlim =     0.000
 radmax =    4.000
 Minimum radius of acceptable spheres?
  1.39999998
 output to  5coxrec.sph
 clustering is complete      49  clusters
```

File 3 is a partial list containing the cluster number and the number of spheres in that particular cluster. The first cluster is usually the largest cluster and is ranked according to the number of spheres in it. In File 3, the cluster number is one and has 314 spheres in it. Figure 14 shows all the 49 clusters on the receptor structure in pdb format.

File 3. The 5coxrec.sph file with co-ordinates of spheres in clusters (a partial file).

```
DOCK 3.5 receptor_spheres
cluster     1   number of spheres in cluster   314
    1  33.86127   5.24224 -20.35150   2.536  126 0  0
    2  37.01283   6.03600 -17.87776   2.555  121 0  0
    6  38.22809  10.20790 -14.99500   3.544  696 0  0
    8  37.48247  10.76998 -13.64997   3.008  694 0  0
   11  32.10894   9.69918  -8.68469   2.682  846 0  0
   12  31.46670   8.01977  -7.57835   1.628   32 0  0
   13  34.59720  10.21953  -9.25875   2.828  684 0  0
   16  31.10007   9.94951 -10.27780   1.955  857 0  0
   28  32.10211   6.51510  -4.09244   2.211   46 0  0
   31  31.37568   7.53117  -3.17091   2.509   45 0  0
   32  32.19233   7.88500  -4.49252   2.557  841 0  0
   33  29.84737   7.92419  -2.58200   2.189  834 0  0
   34  31.09390  10.21620  -8.27045   2.173  846 0  0
   38  29.60381  -2.62915  -2.16924   2.443  202 0  0
   43  31.71706   5.35851  -4.30791   1.436   29 0  0
   45  29.81698   7.38218   0.14289   2.786 2054 0  0
   46  31.76289   5.66931  -4.19217   1.628   29 0  0
   49  29.05761   5.99802   0.99937   2.243 2083 0  0
   50  29.25646   5.99636   1.38381   2.247 2083 0  0
   52  29.02018  -3.41387  -1.70898   3.030  202 0  0
   56  28.82901   6.30148  -0.89333   1.758 2054 0  0
   57  27.41262   3.52551   4.01783   2.298 2083 0  0
   58  27.61276   3.92113   4.48495   2.219 2083 0  0
   59  30.15703   5.77092   5.98702   2.106 2082 0  0
   60  32.97385   3.12507   7.13479   1.535  197 0  0
```

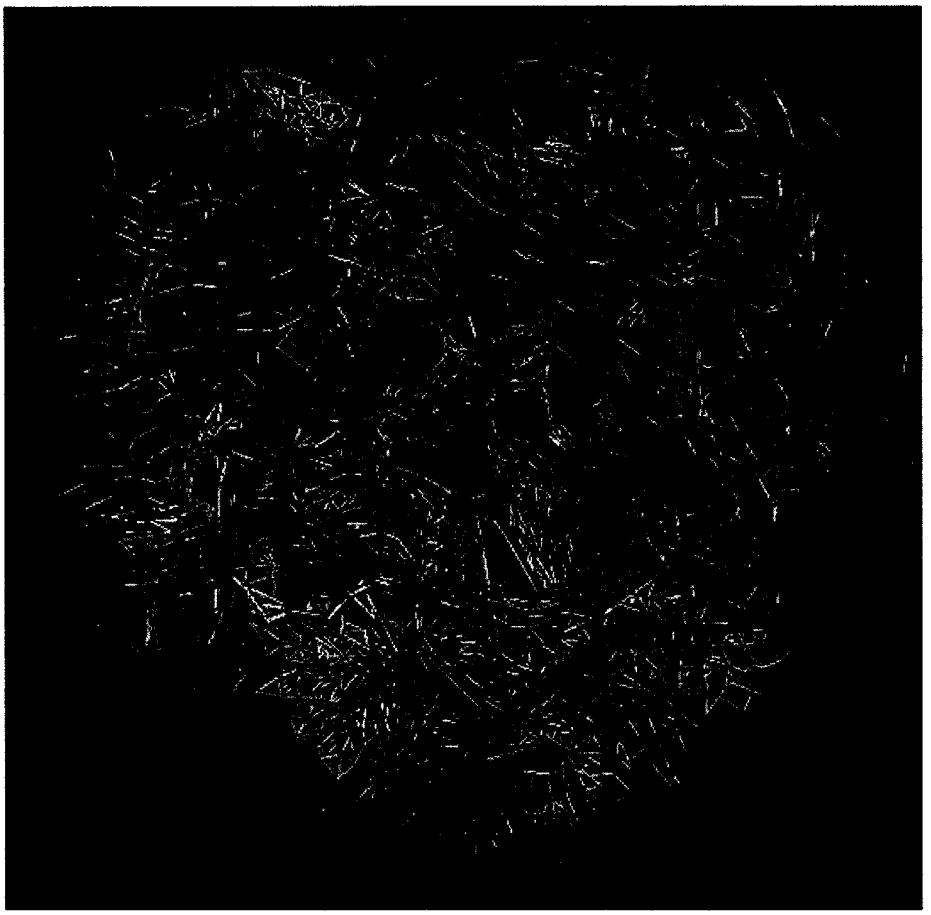


Figure 14. The 5coxrec.sph file in pdb format.

The 5coxrec.sph file was edited by splitting into 49 different cluster files. Then the cluster files were merged with the modified receptor file (5coxrec_noH.pdb) individually. The VEGA ZZ program was used to check if the spheres in the cluster were resting on the active site. It was found that the spheres in 5th cluster file matched with the active site of the receptor. The fifth cluster file is shown in File 4.

File 4. The selected fifth cluster file.

```
cluster     5   number of spheres in cluster    28
  564  25.47775  21.24536  18.13414   2.069 1440 0  0
 1407  24.89094  23.01594  15.34170   2.629 2382 0  0
 1408  25.53015  23.88636  14.41222   2.687 2379 0  0
 1432  24.32610  24.15849  16.15387   1.964 1407 0  0
 1433  24.53161  23.30254  14.48500   2.882 2356 0  0
 1435  24.48589  23.48287  13.64364   2.930 2356 0  0
 1440  24.55522  21.06678  17.15050   2.280  564 0  0
 1455  26.13715  21.23123  19.92801   1.536 1494 0  0
 1456  24.61458  20.51658  17.57876   1.950  564 0  0
 1494  26.41505  21.02241  19.82538   1.653 1455 0  0
 1713  25.27725  24.35810  10.50218   1.669 1724 0  0
 1724  25.48431  24.41288  11.92778   2.226 2376 0  0
 1748  24.78060  23.99746  12.08374   2.347 2375 0  0
 1750  24.86185  24.11027  12.70872   2.652 2376 0  0
 2320  22.65622  23.47803  14.64984   1.740 1431 0  0
 2345  23.89269  23.47197  11.35615   2.015 1748 0  0
 2351  24.74022  22.05213  13.03681   1.837 2378 0  0
 2355  24.42115  21.02557  17.17635   2.223 1440 0  0
 2356  24.56858  23.41748  13.93200   2.932 2379 0  0
 2375  24.12689  23.52305  11.48329   2.140 1748 0  0
 2376  24.56398  23.84492  12.97488   2.786 1748 0  0
 2378  24.70856  21.80117  12.29410   1.569 2351 0  0
 2379  24.87683  23.63687  14.30022   2.894 1435 0  0
 2382  24.49579  22.54498  15.31068   2.528 2356 0  0
 2405  27.95099  23.52323  15.73837   1.740 1408 0  0
 2412  28.46263  22.79210  15.92203   1.688 2379 0  0
 2413  27.56845  21.04100  18.21653   1.837  564 0  0
 2414  29.08758  23.93856  16.37423   1.428 1408 0  0
```

Figure 15 shows the spheres of the cluster file merging with the active site of the receptor file. The fifth cluster contains 28 spheres shown as solid spheres. The tubular structures are valine 523 (yellow) and phenylalanine 518 (purple) residues which are near the active site.

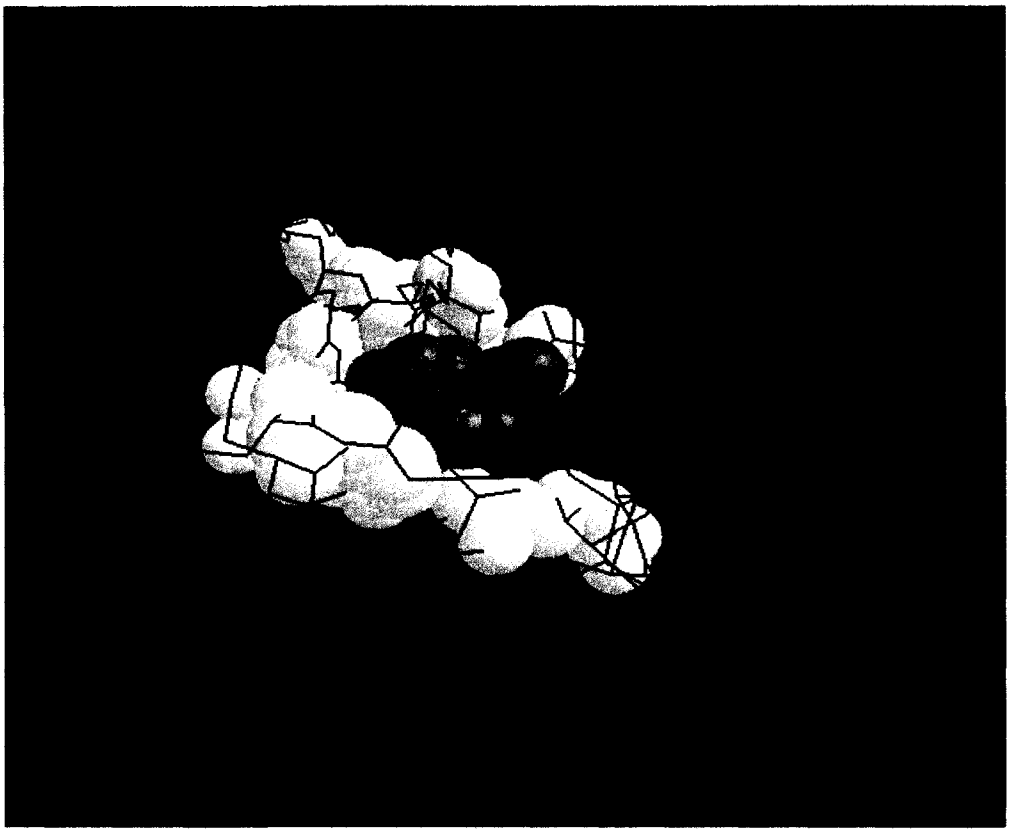


Figure 15. Green spheres filling up the active site of the target structure.

This step demands a lot of visual inspection and time. If a different cluster file is chosen, it leads to a different output during the final docking step, sometimes resulting in an error message during docking. Previous studies have shown that the active site surrounded by tyrosine385, tryptophan 387, phenylalanine 518, tyrosine 248, leucine 352, arginine 120, glutamic acid 524, tyrosine 355, valine 523, and phenylalanine 518 residues. Based on this information, the residues were marked in white using the VEGA ZZ program and it was determined if the spheres were resting on the active site.

The exact cluster file will be given to the students after explaining the procedure for cluster selection. From the students' point of view, this will help them save time and avoid errors in cluster selection.

6.5.4 Scoring Grids for the Target

The purpose of this step is to generate energy scoring grids for the target structure. Prior to grid generation a box is constructed around the active site. The box specifies the location and size of the grid to be calculated. To create a box around the active site, a program called Show Box is used. The input file 5coxbox.in contains default parameters, the cluster file name, and the output file name. Figure 16 shows the output file 5_5coxrec_box.pdb.

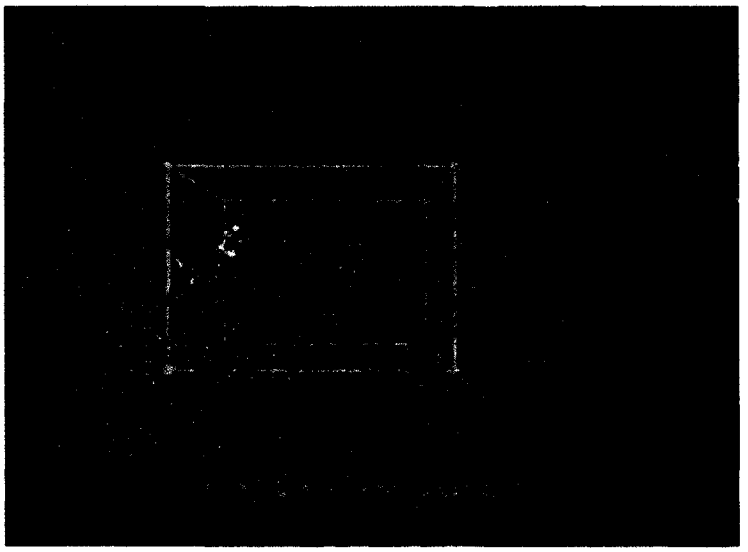


Figure 16. Box enclosing the active site of target and input file.

To generate the scoring grid, a program called Grid is used. This program calculates the force field scores that are approximate molecular mechanics interaction energies, comprised of sum of van der Waals and electrostatic interactions between the ligand and the receptor for the site enclosed by the grid box. The output from the calculation is stored as a scoring grid. The ligand-receptor binding energy is the sum of van der Waals attractive, van der Waals dispersive, and columbic electrostatic energies.

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^{a}} - \frac{B_{ij}}{r_{ij}^{b}} + 332 \frac{q_i q_j}{D r_{ij}} \right) \qquad \text{Equation 2}$$

- E is the intermolecular interaction energy in Kcal/mol
- i and j are the ligand and the receptor atoms
- a and b are the van der Waals repulsive and attractive exponents
- Aij and Bij are the van der Waals repulsion and attraction parameters
- r_{ij} is the distance between atoms i and j
- q_i and q_j is the point charges on the ligand i and receptor j atoms
- D is the dielectric function
- 332 is the factor to convert electrostatic energy to kcal/mol

The input file 5coxgrid.in is created, consisting of the target file, box file and vdw definition file along with the default parameters (obtained from demo version of DOCK). File 5 is the input file for the grid program.

File 5. The input file 5coxgrid.in for grid program.

```
compute_grids                   yes
grid_spacing                    0.3
output_molecule                 no
contact_score                   no
energy_score                    yes
energy_cutoff_distance          9999
atom_model                      a
attractive_exponent             6
repulsive_exponent              12
distance_dielectric             yes
dielectric_factor               4
bump_filter                     yes
bump_overlap                    0.75
receptor_file                   5coxrec_charged.mol2
box_file                        5_5coxrec_box.pdb
vdw_definition_file             vdw_AMBER_parm99.defn
score_grid_prefix               grid
```

The command grid -i 5coxgrid.in is run. Grid generates two binary files, namely grid.bmp and grid.nrg. The grid.bmp file is a filter to remove ligand atoms that overlap on the target atoms during docking. The grid.nrg contains scoring grid information.

For the ligand molecule, the structure file in the mol file format is obtained from KEGG Ligand Database and is opened and saved as ligandA.mol2 in VEGA ZZ. Using VEGA ZZ hydrogen atoms are added to complete the valency; charges are calculated for the ligand and are saved as ligandAcharged.mol2. The structures of ligands are given to the students and ligands' name remains unknown. The chemical names of the ligands are listed below. Figure 17 shows the chemical structures of the ligands in study.

a. Ligand A = Acetyl salicylic acid
b. Ligand B = Rofecoxib
c. Ligand C = Celecoxib
d. Ligand D = SC-558 (1-Phenylsulfonamide-3-trifluoromethyl-5-parabromophenylpyrazole)
e. Ligand E=Fucoxanthin

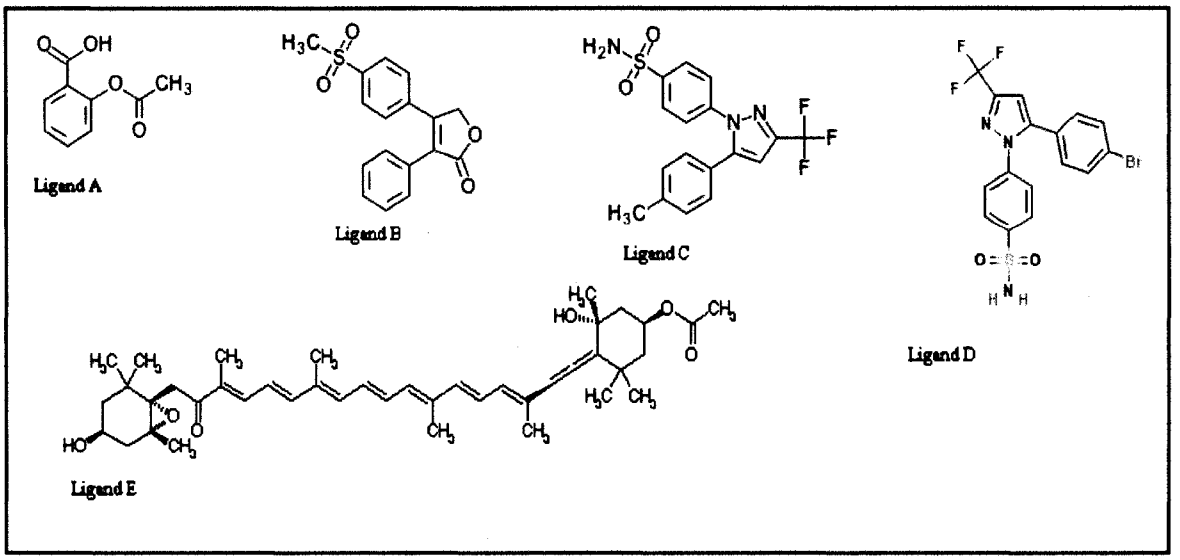


Figure 17. The chemical structures of Ligand A, B, C, D, and E.

Figure 18 is the structure of sample ligand acetyl salicylic acid followed by Figure 19 showing the ligand with hydrogen atoms added to it.

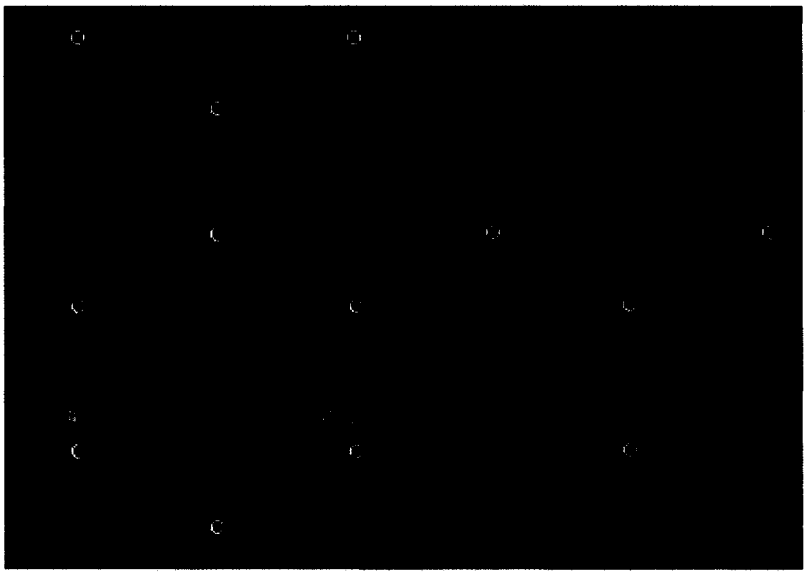


Figure 18. The structure of Ligand A (acetyl salicylic acid).

Figure 19. Hydrogen atoms are added to the Ligand A.

6.5.5 Docking the Target and Ligand

The purpose of this step is to bind the ligand to the target at the site chosen (active site) in the best orientation possible. The files ligandAcharged.mol2, cluster file (5cox_selected spheres.sph), and the grid file (grid.nrg) were included in the input file called 5coxanchor_and_grow.in. The default parameters from the input file of the demo version of DOCK were also added. The input file 5coxanchor_and _grow.in is shown in File 6.

File 6. The input file 5coxanchor_and_grow.in. for Ligand A.

```
ligand_atom_file                                ligandAcharged.mol2
ligand_outfile_prefix                                          flex
limit_max_ligands                                              no
read_mol_solvation                                             no
write_orientations                                             no
write_conformations                                            no
skip_molecule                                                  no
calculate_rmsd                                                 no
rank_ligands                                                   no
num_scored_conformers_written                                  1
orient_ligand                                                  yes
automated_matching                                             yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                               600
critical_points                                                no
chemical_matching
use_ligand_spheres                                             no
flexible_ligand                                                yes
min_anchor_size                                                40
num_anchor_orients_for_growth                                  100
number_confs_for_next_growth                                   100
use_internal_energy                                            yes
internal_energy_att_exp                                        6
internal_energy_rep_exp                                        12
internal_energy_dielectric                                     4.0
use_clash_overlap                                              no
bump_filter                                                    no
score_molecules                                                yes
contact_score_primary                                          no
contact_score_secondary                                        no
grid_score_primary                                             yes
grid_score_secondary                                           yes
grid_score_vdw_scale                                           1
grid_score_es_scale                                            1
grid_score_grid_prefix                                         grid
minimize_ligand                                                yes
minimize_anchor                                                yes
minimize_flexible_growth                                       yes
use_advanced_simplex_parameters                                no
simplex_max_cycles                                             1
simplex_score_converge                                         0.1
simplex_cycle_converge                                         1.0
simplex_trans_step                                             1.0
simplex_rot_step                                               0.1
```

```
simplex_tors_step                                    10.0
simplex_anchor_max_iterations                        500
simplex_grow_max_iterations                          500
simplex_final_min_add_internal                       no
simplex_secondary_minimize_pose                      no
simplex_random_seed                                  0
atom_model                                           all
vdw_defn_file                        vdw_AMBER_parm99.defn
flex_defn_file                                  flex.defn
flex_drive_file                            flex_drive.tbl
```

With the help of File 6 (5coxanchor_and_grow.in), the final stage of docking is performed. Using the Cygwin window, DOCK is run using the command "dock –i dock. in –i dock. out", dock. in is the input file (5coxanchor_and _grow. in) and dock. out is the output file. DOCK uses all the parameters from the input file for docking and if some data is not available it will be requested from the user. The above default parameters could be changed to experiment with the software. The output result is shown in File 7. At the end of the file is the lowest binding energy score of the best conformation for the ligand. The ligand docked into the receptor site is also shown in Figure 20.

File 7. The output file LigandA.docked.out.

```
------------------------------------
DOCK v6.0

Released June 2006
Copyright UCSF
------------------------------------
Molecule Library Parameters
----------------------------------------------------------------------
ligand_atom_file                                ligandAcharged.mol2
ligand_outfile_prefix                                          flex
limit_max_ligands                                                no
read_mol_solvation                                               no
write_orientations                                               no
write_conformations                                              no
skip_molecule                                                    no
calculate_rmsd                                                   no
rank_ligands                                                     no
num_scored_conformers_written                                     1

Orient Ligand Parameters
----------------------------------------------------------------------
orient_ligand                                                   yes
automated_matching                                              yes
receptor_site_file                        5coxselectedspheres.sph
max_orientations                                                600
critical_points                                                  no
chemical_matching                                                no
use_ligand_spheres                                               no

Flexible Ligand Parameters
----------------------------------------------------------------------
flexible_ligand                                                 yes
min_anchor_size                                                  40
num_anchor_orients_for_growth                                   100
number_confs_for_next_growth                                    100
use_internal_energy                                             yes
internal_energy_att_exp                                           6
internal_energy_rep_exp                                          12
internal_energy_dielectric                                      4.0
use_clash_overlap                                                no
Bump Filter Parameters
 bump_filter                                                      no
Master Score Parameters
score_molecules                                                 yes
Contact Score Paramters
----------------------------------------------------------------------
contact_score_primary                                            no
contact_score_secondary                                          no
```

```
Grid Score Parameters
-----------------------------------------------------------------
grid_score_primary                                              yes
grid_score_secondary                                            yes
grid_score_vdw_scale                                            1
grid_score_es_scale                                             1
grid_score_grid_prefix                                          grid
Simplex Minimization Parameters
-----------------------------------------------------------------
minimize_ligand                                                 yes
minimize_anchor                                                 yes
minimize_flexible_growth                                        yes
use_advanced_simplex_parameters                                 no
simplex_max_cycles                                              1
simplex_score_converge                                          0.1
simplex_cycle_converge                                          1.0
simplex_trans_step                                              1.0
simplex_rot_step                                                0.1
simplex_tors_step                                               10.0
simplex_anchor_max_iterations                                   500
simplex_grow_max_iterations                                     500
simplex_final_min_add_internal                                  no
simplex_secondary_minimize_pose                                 no
simplex_random_seed                                             0
Atom Typing Parameters
-----------------------------------------------------------------
atom_model                                                      all
vdw_defn_file
vdw_AMBER_parm99.defn
flex_defn_file
flex.defn
flex_drive_file
flex_drive.tbl
Initializing Library File Routines...
Initializing Orienting Routines...
Initializing Conformer Generator Routines...
Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg
-----------------------------------
Molecule: ligandAcharged
 Elapsed time:    663 seconds

 Anchors:         1
 Orientations:          600
 Conformations:         83

 Grid Score:           -38.733078
          vdw:           -28.639036
           es:           -10.094044
1 Molecules Processed
Total elapsed time:    664 seconds
```

The receptor structure shown in Figure 20 is shown in green and the docked ligand is shown in purple using UCSF Chimera. It is to be noted that DOCK outputs only the best oriented ligand in mol2.format. The receptor structure file was merged with the oriented ligand using the VEGA ZZ program to view the complete output.

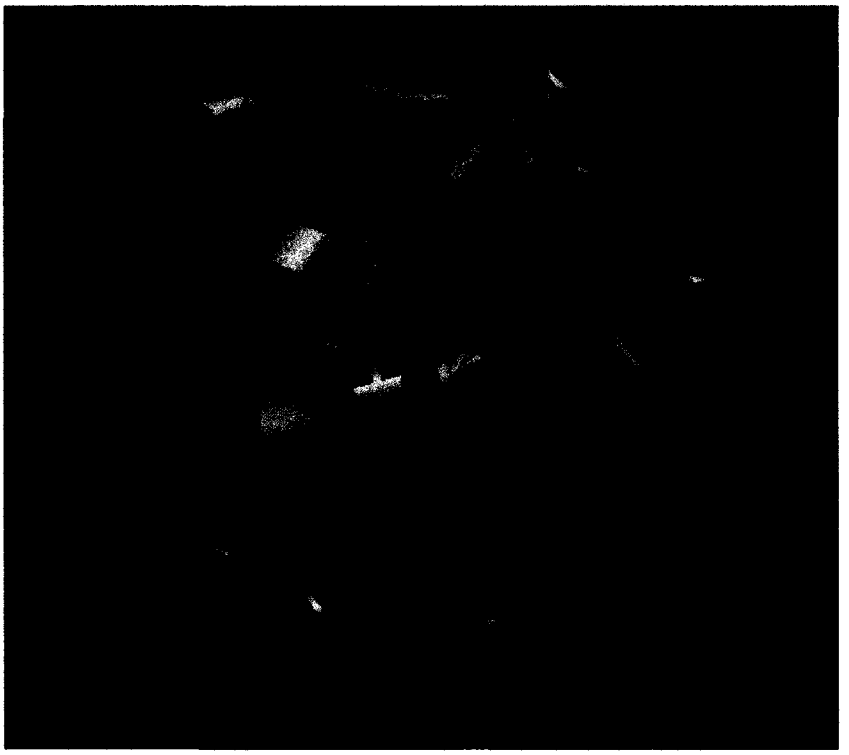


Figure 20. Ligand A docked into red receptor structure in flexible docking.

The results presented in File 7 and Figure 20 is for flexible docking. Another form of docking is called rigid docking in which the receptor and ligand are assumed to be rigid structures and do not change spatially during docking process. The input file for rigid docking is given in File 8. The command for rigid docking is "dock6 –i rigid.in." and DOCK outputs two files rigid.out shown in File 9 and rigid_scored.mol2 shown in Figure 21 generated from UCSF Chimera.

File 8. The input file for rigid docking 5coxrigid.in.

```
ligand_atom_file                                ligandAcharged.mol2
ligand_outfile_prefix                                        rigid
limit_max_ligands                                               no
read_mol_solvation                                              no
write_orientations                                              no
write_conformations                                             no
skip_molecule                                                   no
calculate_rmsd                                                  no
rank_ligands                                                    no
num_scored_conformers_written                                    1
orient_ligand                                                  yes
automated_matching                                             yes
receptor_site_file                  5coxselectedspheres.sph
max_orientations                                              1000
critical_points                                                 no
chemical_matching                                               no
use_ligand_spheres                                              no
flexible_ligand                                                 no
bump_filter                                                     no
score_molecules                                                yes
contact_score_primary                                           no
contact_score_secondary                                         no
grid_score_primary                                             yes
grid_score_secondary                                           yes
grid_score_vdw_scale                                             1
grid_score_es_scale                                              1
grid_score_grid_prefix                                        grid
minimize_ligand                                                yes
simplex_max_iterations                                        1000
simplex_max_cycles                                               1
simplex_score_converge                                         0.1
simplex_cycle_converge                                         1.0
simplex_trans_step                                             1.0
simplex_rot_step                                               0.1
simplex_tors_step                                             10.0
simplex_final_min_add_internal                                  no
simplex_secondary_minimize_pose                                 no
simplex_random_seed                                              0
atom_model                                                     all
vdw_defn_file                     vdw_AMBER_parm99.defn
flex_defn_file                                  flex.defn
flex_drive_file                            flex_drive.tbl
```

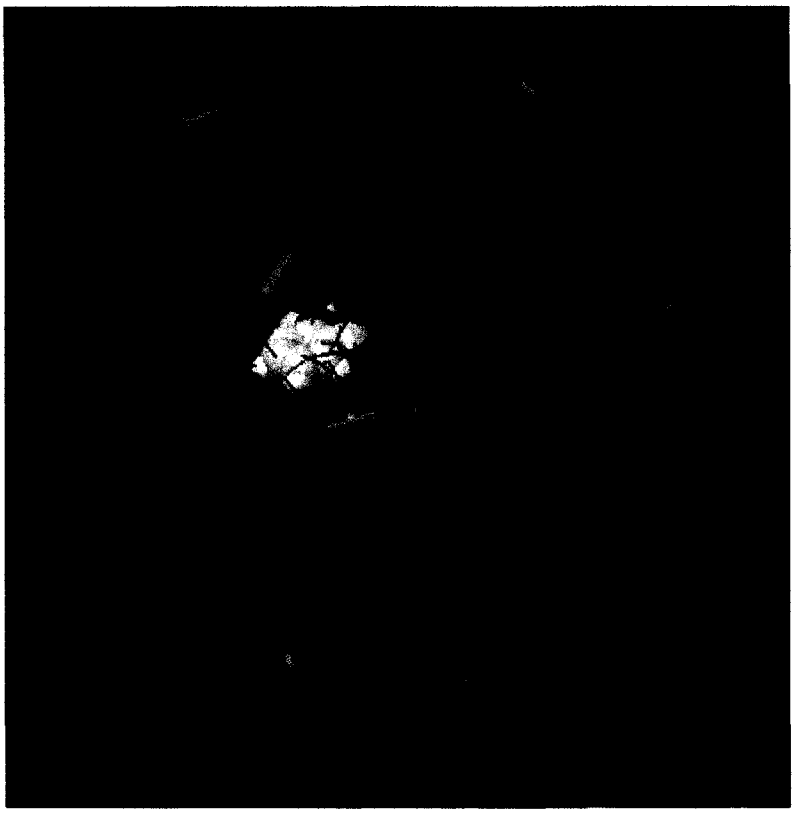


Figure 21. Rigid docking of Ligand A in receptor structure.

File 9. The output file for rigid docking.

```
------------------------------------
DOCK v6.0
Released June 2006
Copyright UCSF
------------------------------------
Molecule Library Parameters
----------------------------------------------------------------
ligand_atom_file                              aspirincharged.mol2
ligand_outfile_prefix                                      rigid
limit_max_ligands                                             no
read_mol_solvation                                            no
write_orientations                                            no
write_conformations                                           no
skip_molecule                                                 no
calculate_rmsd                                                no
rank_ligands                                                  no
num_scored_conformers_written                                  1
```

```
Orient Ligand Parameters
-------------------------------------------------------------------
orient_ligand                                                   yes
automated_matching                                              yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                               1000
critical_points                                                 no
chemical_matching                                               no
use_ligand_spheres                                              no
Flexible Ligand Parameters
flexible_ligand                                                 no
Bump Filter Parameters
bump_filter                                                     no
Master Score Parameters
-------------------------------------------------------------------
score_molecules                                                 yes
Contact Score Paramters
contact_score_primary                                           no
contact_score_secondary                                         no
 Grid Score Parameters
 grid_score_primary                                              yes
 grid_score_secondary                                            yes
 grid_score_vdw_scale                                            1
 grid_score_es_scale                                             1
 grid_score_grid_prefix
 grid
 Simplex Minimization Parameters
 -------------------------------------------------------------------
 minimize_ligand                                                 yes
 simplex_max_iterations                                         1000
 simplex_max_cycles                                              1
 simplex_score_converge                                          0.1
 simplex_cycle_converge                                          1.0
 simplex_trans_step                                              1.0
 simplex_rot_step                                                0.1
 simplex_tors_step                                              10.0
 simplex_final_min_add_internal                                  no
 simplex_secondary_minimize_pose                                 no
 simplex_random_seed                                             0
 Atom Typing Parameters
 -------------------------------------------------------------------
 atom_model                                                      all
 vdw_defn_file                          vdw_AMBER_parm99.defn
 flex_defn_file                                 flex.defn
 flex_drive_file                          flex_drive.tbl
 Initializing Library File Routines...
 Initializing Orienting Routines...
 Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg
 Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg
```

```
---------------------------------
Molecule: ligandAcharged
Elapsed time:    69 seconds

Anchors:         1
Orientations:          1000
Conformations:         1000

Grid Score:           -36.690540
          vdw:          -28.418787
           es:           -8.271753
1 Molecules Processed
Total elapsed time:    69 seconds
```

The results for the other ligands are presented in Appendix A. The energy scores have been tabulated in Table 1 for rigid and flexible docking. The more negative the score, the higher the binding affinity. The computer output of these numbers have been expressed to three significant figures in light of accuracy of the data, the calculations are based. Figure 22 and Figure 23 shows an ensemble of all the ligands in various colors docked in the receptor binding site.

Table 1. The binding scores of the docked ligands.

Ligand	Grid score for flexible docking Kcal/mol	Grid score for rigid docking Kcal/mol
Ligand A(acetyl salicylic acid)	-38.7	-36.6
Ligand B(rofecoxib)	-40.3	-35.4
Ligand C(celecoxib)	-34.5	-14.5
Ligand D(SC-558)	- 40.7	13.4
Ligand E(fucoxanthin)	Error: Could not complete growth. Confirm grid box is large enough to contain ligand	Error: Conformation could not be scored by DOCK. Conformation not completely within grid box.

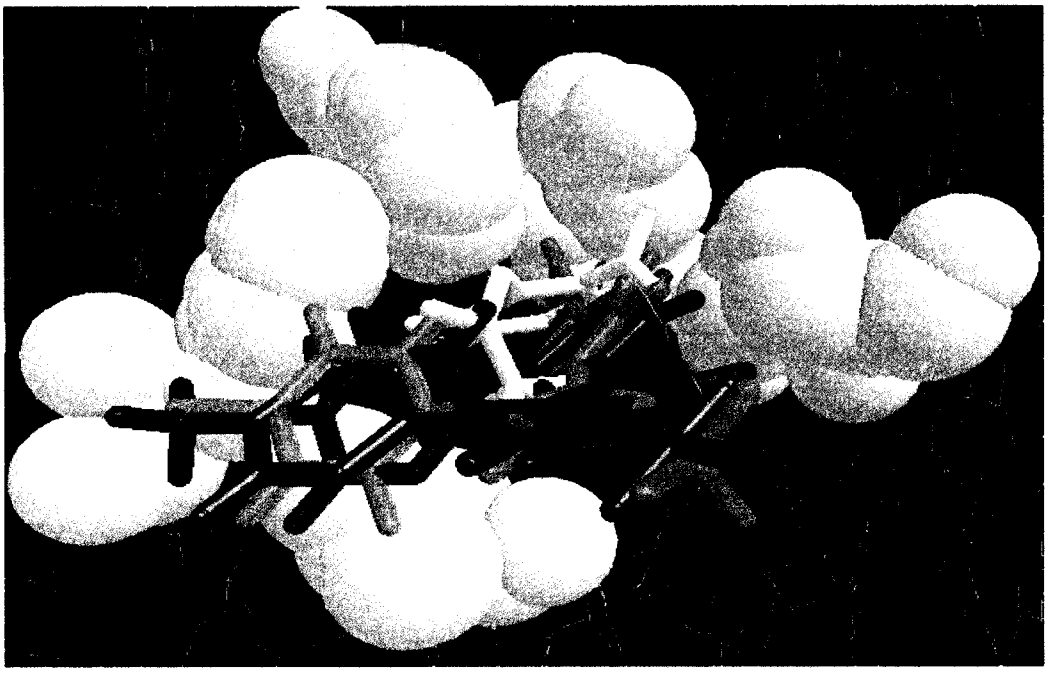


Figure 22. Three ligands A, B, C, and D bound to receptor in flexible docking.

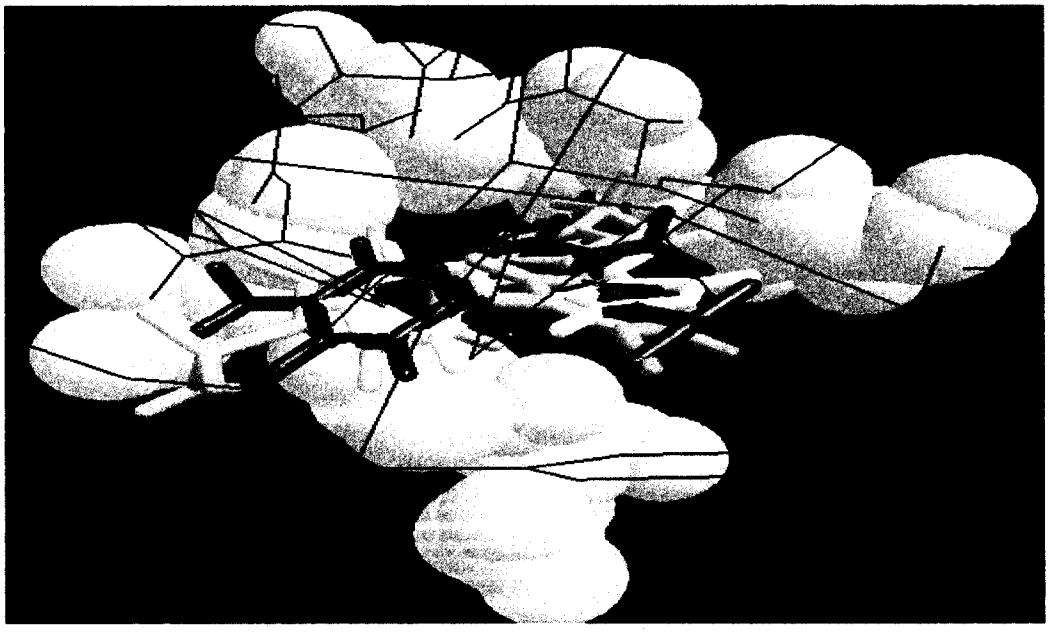


Figure 23. Four ligands A, B, C, and D bound to the receptor in rigid docking.

The results from Table 1 for flexible docking show that Ligand D (SC-558) and Ligand B (rofecoxib) have the lowest binding scores -40.7 and -40.3 kcal/mol, indicating high binding affinity. This is followed by Ligand A (acetyl salicylic acid) with a binding score of -38.7 kcal/mol. Ligand C (celecoxib), when compared with the other ligands has the least binding score (-34.5 kcal/mol) exhibiting less binding interactions inspite of the fact that like Ligands B and D, it belongs to the coxib group. Ligand B, Ligand C and Ligand D belong to the diaryl heterocyclic group that has a central pyrazole ring and a sulphonamide substituent attached to the aryl ring. The sulphonamide group of ligands B, C, and D binds into the side pocket of the COX-2 active site as shown in Figure 23. Ligand E could not be docked because the DOCK software gave an error message stating that the grid box enclosing the active site was too small to fit the ligand. Increasing the gridbox had no effect on the result. Based on the flexible docking results, all the ligands except Ligand E can be considered as potential lead candidates.

Rigid docking is considered a basic mode of docking because the receptor and the ligand are assumed to be non-flexible. Ligand A (acetyl salicylic acid: -36.6 kcal /mol) has the lowest energy score resulting in high binding affinity and is closely followed by Ligand B (-35.400 kcal/mol). Ligand C (celecoxib) has a lowest binding score of -14.5 kcal mol displaying least binding affinity. Ligand D (SC-558) exhibited a positive score resulting in no binding affinity. The Ligands B, C, and D belong to the same family (the coxib group), yet Ligands D and C display poor binding affinity-this could be due to an additional functional group. Ligand E (fucoxanthin) could not be docked because the conformation was too large to fit inside the grid box. This clearly shows that Ligand E is

too large a ligand to fit inside the COX-2 active site. Based on rigid docking results, Ligand A and Ligand B alone can be considered good lead compounds.

6.6 Ligand Evaluation

Five ligands were docked into the receptor structure individually using the DOCK software. The ligands are evaluated based on Lipinski's Rule of Five which states that poor absorption or permeation is more likely when:

- There are more than five H-bond donors.
- The molecular weight is over 500.
- The logP (partition coefficient) is over 5.
- The sums of N's and O's are over 10.

Based on Veber's paper, the number of rotatable bonds should be less than 10 for oral bioavailability. Based on the five characteristics, the ligands are tabulated in Table 2 based on values obtained from Pubmed database.

Table 2. Ligand with "Lipinski Rule of Five" characteristics.

Ligand	LogP Partition coefficient	OH+NH H-bond donors	Mol Wt g/mol	N+O H-bond acceptors	Rotatable bonds	Alert 1: good; 0: poor
Ligand A(acetyl salicylic acid)	1.4	1	180.16	4	3	1
Ligand B (rofecoxib)	3.2	0	314.357	4	3	1
Ligand C (celecoxib)	3.9	1	381.373	8	3	1
Ligand D (SC-558)	4.2	1	446.243	8	3	1
Ligand E(fucoxanthin)	7.9	2	658.906	6	12	0

Examination of Table 2 shows that ligands A-D obey the Rule of 5 to a greater or lesser extent. Ligand E does not obey the criteria and can be considered a poor ligand. Ligand D is at the borderline; all the other ligands are well within the limit. It is known that permeability decreases with high molecular weight, resulting in less bioabsorption [35]. Log P is the partition coefficient; this parameter is found to be directly proportional to the molecular weight. As the molecular weight increases, the log P values become larger, indicating lower permeability shown in Figure 24.

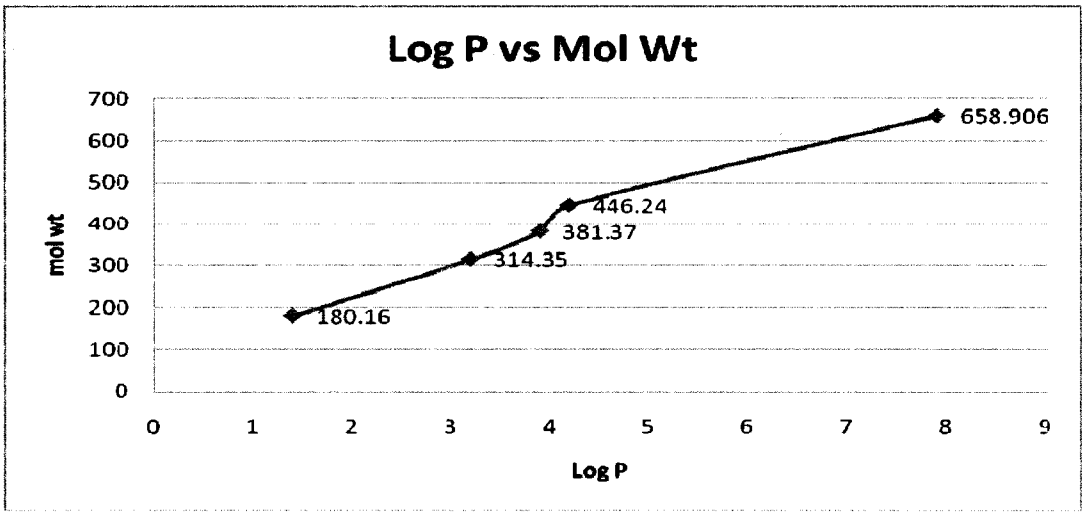


Figure 24. Graph showing LogP vs. molecular weight.

Comparing the results obtained from DOCK and those of Lipinski's Rule of Five, Ligand A is the smallest compound, and it fits into the active site like a ball in a glove. The enclosed binding site allows enough space for Ligand A to move inside the site, thereby displaying a low energy score. The energy score from rigid docking is similar to that of flexible docking (-38.733 kcal/mol). Also, Ligand A penetrates deep into the hydrophobic pocket showing that it can also bind to the COX-1 active site in a similar fashion (49). This shows that it is a non- selective inhibitor and that there is a possibility that it will inhibit both COX-1 and COX-2 when it enters the human body. Our aim is to inhibit only COX-2, so Ligand A does not serve the purpose of a good lead in spite of its good binding score. Ligand B (rofecoxib) binds to the side pocket of the COX-2 active site in rigid and flexible docking. The side pocket feature is unique only to COX-2 and there is less probability that Ligand B would bind to COX-1 when it enters in the human system. The same argument can be applied to Ligands B, C, and D (coxib group) as they

all bound to the side pocket of COX-2 and showed high binding affinity. They can be good lead compounds as well as selective inhibitors for COX-2. Ligand D has a bromine atom attached to one of the rings; it might be eliminated during the preclinical testing phase due to the toxic properties of the halogen. The ligand-enzyme complexes for Ligands B and C (available from literature) are compared with the results obtained in this study-this is presented in Appendix A.

The binding energies of these ligands need to be determined using experimental methods and compared with the predicted results. If the RMSD values of the predicted and the experimental values are similar, then the ligand could be further studied for preclinical testing.

CHAPTER SEVEN

CONCLUSIONS

The process of structure-based drug design was studied. The drug design concepts were applied to a case study that finds the top scoring ligand using COX-2 as a drug target. The COX-2 enzyme is an excellent drug target for osteoarthritis. The UCSF DOCK program was used to predict the binding energies and the VEGA ZZ program to visualize the results. Five known ligands were treated as novel compounds for the drug design study. It was found that Ligand B (rofecoxib), Ligand C (celecoxib), and Ligand D (SC-558) were good inhibitors of COX-2 based on docked results. Based on Lipinski's Rule of Five, it was found that the permeation of the ligand depends on the molecular weight and the partition coefficient. To prove the real potential of the ligands used in this case study, they need to be tested experimentally.

There is always a possibility that the predicted values may not be equal to those of the experimental values. It is probable that the lower scoring compounds might perform better under experimental conditions. Apart from structural details, factors such as bioabsorption, chemical stability, and production costs also play an important role in determining the lead drug.

REFERENCES

1. The National Institute of General Medical Sciences (2004) NIGMS-Supported Structure-based Drug Design Saves Lives [online] Available at http://www.nigms.nih.gov/news/facts/structure_drugs.html (accessed September 24, 2004).

2. University of Waterloo (2004) Drug discovery in the bioinformatics age [online] Available at http://monod.uwaterloo.ca/cs798/proteomics.pdf (accessed September 24, 2004).

3. The Wellcome Trust (2004) An overview of drug development [online] Available at http://www.wellcome.ac.uk/en/genome/tacklingdisease/hg09b005.html (accessed September 24, 2004).

4. The National Institute of General Medical Sciences (2004) NIGMS-Supported Structure-based Drug Design Saves Lives [online] Available at http://publications.nigms.nih.gov/basicresearch/ (accessed September 24, 2004).

5. Graham H. Coombs, Daniel E. Goldberg, Michael Klemba, Colin Berry, John Kay and Jeremy C. Mottram , " *Aspartic proteases of Plasmodium falciparum and other parasitic protozoa as drug targets,"* Trends in Parasitology **17(11),** 2001.

6. R. Banerjee, S.E. Francis and D.E. Goldberg, "*Food vacuole plasmepsins are processed at a conserved site by an acidic convertase activity in Plasmodium falciparum*, "*Mol. Biochem. Parasitol.* **129**,157–165 (2003).

7. Kishore R. Sakharkar, Meena K. Sakharkar and Vincent T.K. Chow, *" A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa,"* in silico Biology, **4**, 0028 (2004).

8. Pearl L.H, Taylor W.R, *"A structural model for the retroviral proteases,"* Nature **329**, 351–354 (1987).

9. Navia M.A, Fitzgerald P.M, McKeever B.M, Leu C.T, Heimbach J.C,"*Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1,"* Nature **337**, 615–20 (1989).

10. Rational drug design approach to synthesizing HIV-Protease inhibitors (Online) Available at http://biochem118.stanford.edu/Projects/1999/ding.pdf (accessed June 2006).

11. National Cancer Institute, Molecular crystallography laboratory [online] Available at http://xpdb.nist.gov/hivsdb/gallery.html (accessed October 14, 2004).

12. Akihiko Kuniyasu, *" Identification of the Ligand Binding Sites and Novel Drug Target Molecules by Photoaffinity Labeling,"* **123(8)**, 673-679 (2003).

13. A. Filikov, V. Mohan, T. Vickers, R. Griffey, P. Cook, R. Abagyan and T. James, *"Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR,"* J.Comput.Aided Mol Des. **6,** 593-610 (2000).

14. M. Hendlich, F. Rippmann and G. Barnickel, *"LIGSITE: automatic and efficient detection of potential small-molecule binding sites in proteins,"* J.Mol.Graph.Model **15,** 359–363 (1997).

15. K. P.Peters, J. Fauck and C. Frommel, *"The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria,"* J.Mol. Biol. **256**, 201–213 (1996).

16. G.P. Brady, Jr and P.F. Stouten, *"Fast prediction and visualization of protein binding pockets with PASS,"* J.Comput.Aided Mol.Des. **14**, 383–401 (2000).

17. LIGSITE: M. Hendlich, F. Rippman and G. Barnickel, SITE2SPHERES [online] Available at http://www.daylight.com/meetings/mug01/Dixon/ligsitedescr.html (accessed October 24, 2004).

18. H. Boehm, *"The computer program, LUDI: a new method for the de novo design of enzyme inhibitors,"* J.Comput.Aided Mol.Des. **6**, 61-78 (1992).

19. H. Boehm, *"Towards the automatic design of synthetically accessible protein ligands: Peptides, amides and peptidomimetics,"* J.Comput.Aided Mol.Des. **10**, 265–272 (1996).

20. Gillet P. Johnson, P. Mata, S. Sike and P. Williams, *"Sprout: a program for structure generation."* J.Comput.Aided Mol. Des. **7**, 127–153 (1993).

21. M. Eisen, D. Wiley, M. Karplus and R. Hubbard, *"Hook: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of macromolecule binding sites"*, Proteins **19,** 199–221 (1994).

22. R. Bohacek and C. McMartin, *"Multiple highly diverse structues complementary to enzyme binding sites: Results of extensive application of a de novo design method incorporating combinatorial growth"*, J. Am. Chem. Soc. **116**, 5560–5571 (1994).

23. P. Goodford, *"A computational procedure for determining energetically favorable binding sites on biologically important macromolecules"*, J. Med. Chem. **28**. 849–857 (1985).

24. P. Bartlett, G. Shea, S. Telfer and S. Waterman, "Caveat: a program to facilitate the structure-derived design of biologically active molecules", *Molecular Recognition: Chemical and Biological Problems*, Royal Society of Cambridge, Cambridge 182–196 (1989).

25. Y. Nishibata and A. Itai, "*Confirmation of usefulness of a structure construction program based on three-dimensional receptor structure for rational lead generation*", J. Med. Chem. **36**, 2921–2928 (1993).

26. P.A. Kollman "*Free Energy Calculations: Applications to Chemical and Biochemical Phenomena.*" Chem. Rev. 93, 2395-2417 (1993).

27. R.L. DesJarlais, G.L. Seibel, I.D. Kuntz, P. S. Furth, J.C. Alvarez, P.R Ortiz de Montellano, D.L. DeCamp, L.M. Babé, and C.S Craik, "*Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease,*" Proc Natl Acad Sci U S A. **87**, 6644–6648 (1990).

28. D. Lorber and B. Shoichet, "*Flexible ligand docking using conformational ensembles.*" Protein Sci. **7**, 938–950 (1998).

29. T.Ewing, S.Makino, G. Skillman and I. Kuntz, "*Dock 4. 0: search strategies for automated molecular docking of flexible molecule databases,*" J. Comput. Aided Mol. Des. **15,** 411–428 (2001).

30. C. Lipinski, F. Lombardo, B.Dominy and P. Feeney, "*Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings,*" Adv. Drug Deliv. Rev. **23,** 3-4 (1997).

31. D. Veber, S. Johnson, H. Cheng, B. Smith, K.Ward and K. Kopple, "*Molecular properties that influence the oral bioavailability of drug candidates,*" J. Med. Chem. **45**, 2615–2623 (2002).

32. Sokoloff, Leon. The biology of degenerative joint disease (Chicago, University of Chicago Press [1969]

33. Giuseppe Ermondia, Giulia Carona, Raelene Lawrenceb, and Dario Longo, "*Docking studies on NSAID/COX-2 isozyme complexes using Contact statistics analysis,*" Journal of Computer-Aided Molecular Design. **18,** 683–696 (2004).

34. Kurumbail R.G, Stevens A.M, Gierse J.K, McDonald J.J, Stegeman R.A, Pak J.Y, Gildehaus D, Miyashiro J.M, Penning T.D, Seibert K, Isakson P.C,and Stallings W.C, " *Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents,*" Nature. **384 (6610)**, 644-8 (1996).

35. The Official UCSF DOCK Web-site [online] Author Therese Lang Available at http://dock.compbio.ucsf.edu/DOCK_6/tutorials/sphere_generation/generating_spheres.htm.
36. The Official UCSF DOCK Web-site [online] Author Therese Lang et al. Available at http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm
37. The Official UCSF DOCK Web-site [online] Author http://dock.compbio.ucsf.edu/DOCK_6/tutorials/grid_generation/generating_grid.html
38. The Official UCSF DOCK Web-site [online] Author Therese Lang Available at http://dock.compbio.ucsf.edu/DOCK_6/tutorials/struct_prep/prepping_molecules.htm
39. Bio computing and drug design [online] Available at http://www. techfak. uni-bielefeld. de/bcd/ForAll/Introd/drugdesign. html (accessed October 14, 2004).
40. Health and Medical Information produced by Doctors [online] Available at http://www. medicinenet.com/rheumatoid%20arthritis/article.htm (accessed October 14, 2004).
41. M. Huynen, Y. Diaz-Lazcoz, and P. Bork, *"Differential genome display. Trends Genet,"* 13, 389-390 (1997).
42. Amy C. Anderson, *"The Process of Structure-Based Drug Design,"* Chemistry & Biology, **10**, 787–797 (2003).
43. UCSF DOCK dock-fans -- Molecular docking and virtual screening discussion group [online] Available at http://blur.compbio.ucsf.edu/mailman/listinfo/dock-fans.

APPENDIX A

LIGAND AND TARGET STRUCTURE INFORMATION AND DOCKING RESULTS

1. Ligand A

The information for Ligand A was obtained from KEGG Ligand Database. The mol file is saved in mol2 format and visualized in VEGA. The information is shown in Figure 1 and mol file for Ligand A is shown in File 1.

KEGG **COMPOUND: C01405**

Help

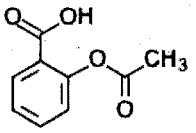
Entry	C01405 Compound
Name	Aspirin; Acetylsalicylic acid; 2-Acetoxybenzenecarboxylic acid; Acetylsalicylate
Formula	C9H8O4
Mass	180.0423
Structure	 C01405 Mol file KCF file DB search
Remark	Same as: D00109
Reaction	R02942
Enzyme	1.14.99.1 (I) 3.1.1.55
Other DBs	CAS: 50-78-2 PubChem: 4594 ChEBI: 15365 3DMET: B00284
LinkDB	All DBs
KCF data	Show

Figure 1. Information for Ligand A in the KEGG Ligand Database.

File 1. The mol file for Ligand A.

```
  13 13  0  0  0  0  0  0  0  0999 V2000
   20.2981  -15.8105    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   21.5226  -16.5029    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   19.0928  -16.5029    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   20.2981  -14.6927    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   21.5226  -17.9133    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   22.7278  -15.8040    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   19.0928  -17.9133    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   21.5033  -13.9940    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   19.0863  -14.0004    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   20.2981  -18.6250    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   23.9396  -16.4964    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   25.1450  -15.7977    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   23.9396  -17.9642    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0     0  0
  1  3  2  0     0  0
  1  4  1  0     0  0
  2  5  2  0     0  0
  2  6  1  0     0  0
  3  7  1  0     0  0
  4  8  1  0     0  0
  4  9  2  0     0  0
  5 10  1  0     0  0
  6 11  1  0     0  0
 11 12  1  0     0  0
 11 13  2  0     0  0
  7 10  2  0     0  0
M  END
```

2. Ligand B

The information for Ligand B was obtained from KEGG Ligand database. The mol file was copied and saved in mol2 format and visualized in VEGA. The information and mol file for Ligand B is shown in Figure 2 and File 2 respectively.

KEGG **COMPOUND: C07590**

Help

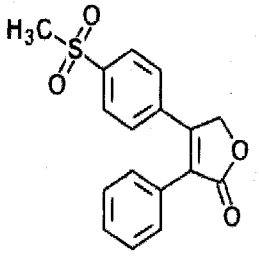
Entry	C07590 Compound
Name	Rofecoxib
Formula	C17H14O4S
Mass	314.0613
Structure	 C07590 Mol file KCF file DB search
Remark	Same as: D00568
Other DBs	CAS: 162011-90-7 PubChem: 9792 ChEBI: 8887
LinkDB	All DBs
KCF data	Show

Figure 2. Information of Ligand B in the KEGG Ligand Database.

File 2. Mol file for Ligand B exported from KEGG Ligand Database.

```
 22 24  0  0  0  0  0  0  0  0999 V2000
    29.9147  -19.5829    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    29.0770  -18.4195    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    27.7739  -18.8384    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    27.7739  -20.2810    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    29.0770  -20.6999    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    29.5424  -22.0030    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    24.1439  -20.9791    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    24.1439  -22.3753    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    25.3539  -23.0734    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    26.5638  -22.3753    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    26.5638  -20.9791    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    25.3539  -20.2810    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    24.1439  -16.7441    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    24.1439  -18.1403    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    25.3539  -18.8384    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    26.5638  -18.1403    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    26.5638  -16.7441    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    25.3539  -16.0460    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    22.9338  -16.0460    0.0000 S   0  0  0  0  0  0  0  0  0  0  0  0
    23.6319  -14.8359    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    22.1892  -17.2560    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    21.6772  -15.3479    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0     0  0
  2  3  1  0     0  0
  3  4  2  0     0  0
  4  5  1  0     0  0
  1  5  1  0     0  0
  5  6  2  0     0  0
  7  8  2  0     0  0
  8  9  1  0     0  0
  9 10  2  0     0  0
 10 11  1  0     0  0
 11 12  2  0     0  0
  7 12  1  0     0  0
 11  4  1  0     0  0
 13 14  2  0     0  0
 14 15  1  0     0  0
 15 16  2  0     0  0
 16 17  1  0     0  0
 17 18  2  0     0  0
 13 18  1  0     0  0
 16  3  1  0     0  0
 13 19  1  0     0  0
 19 20  2  0     0  0
 19 21  2  0     0  0
 19 22  1  0     0  0
M  END
```

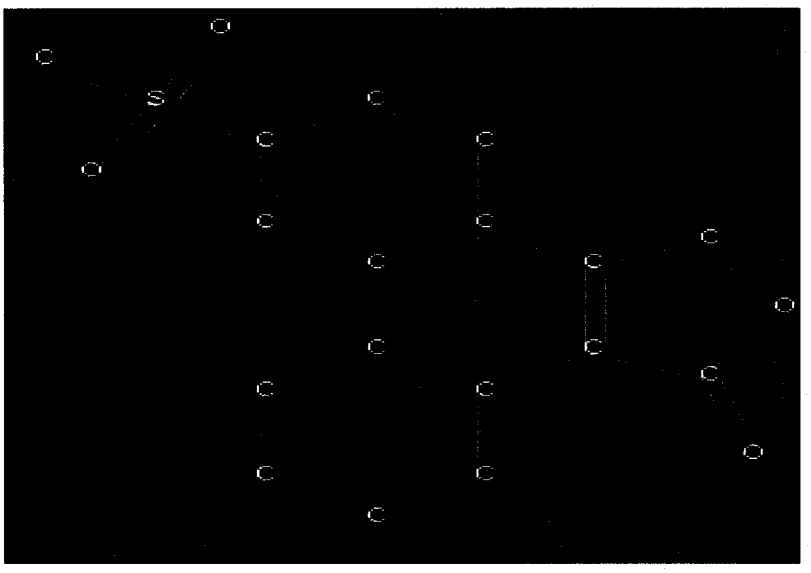


Figure 3. Ligand B chemical structure.

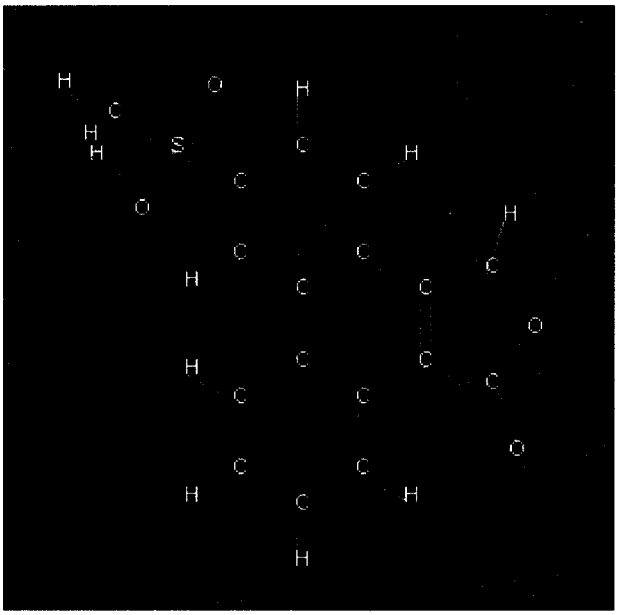


Figure 4. Hydrogen atoms are added to Ligand B.

Hydrogen atoms are added to the ligand file; charges are calculated and saved as ligandBcharged.mol2. The total charge is 0.0 calculated using Gasteiger charges in the VEGA ZZ program.

For flexible docking the files ligandBcharged.mol2, cluster file (5cox_selected spheres.sph), and the grid file (grid. nrg) was incorporated into the input file called 5coxanchor_and_grow. in. The default parameters from input file of demo version of DOCK were also added. The input file 5coxanchor_and _grow. in is shown in File 3.

File 3. Input file 5coxanchor_and _grow.in.

```
ligand_atom_file                                ligandBcharged.mol2
 ligand_outfile_prefix                                         flex
limit_max_ligands                                              no
read_mol_solvation                                             no
write_orientations                                             no
write_conformations                                            no
skip_molecule                                                  no
calculate_rmsd                                                 yes
use_rmsd_reference_mol                                         1
rank_ligands                                                   yes
max_ranked_ligands                                             3
scored_conformer_output_override                               no
orient_ligand                                                  yes
automated_matching                                             yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                               600
critical_points                                                no
chemical_matching                                              no
use_ligand_spheres                                             no
flexible_ligand                                                yes
min_anchor_size                                                40
num_anchor_orients_for_growth                                  100
number_confs_for_next_growth                                   100
use_internal_energy                                            yes
internal_energy_att_exp                                        6
internal_energy_rep_exp                                        12
internal_energy_dielectric                                     4.0
use_clash_overlap                                              no
bump_filter                                                    no
score_molecules                                                yes
contact_score_primary                                          no
contact_score_secondary                                        no
```

```
grid_score_primary                                  yes
grid_score_secondary                                yes
grid_score_vdw_scale                                1
grid_score_es_scale                                 1
grid_score_grid_prefix                              grid
minimize_ligand                                     yes
minimize_anchor                                     yes
minimize_flexible_growth                            yes
use_advanced_simplex_parameters                     no
simplex_max_cycles                                  1
simplex_score_converge                              0.1
simplex_cycle_converge                              1.0
simplex_trans_step                                  1.0
simplex_rot_step                                    0.1
simplex_tors_step                                   10.0
simplex_anchor_max_iterations                       500
simplex_grow_max_iterations                         500
simplex_final_min_add_internal                      no
simplex_secondary_minimize_pose                     no
simplex_random_seed                                 0
atom_model                                          all
vdw_defn_file
vdw_AMBER_parm99.defn
flex_defn_file                                      flex.defn
flex_drive_file
flex_drive.tbl
```

File 4. The output file ligandBdocked.out.

```
DOCK v6.0
Released June 2006
Copyright UCSF
Molecule Library Parameters
-----------------------------------------------------------------------
ligand_atom_file
ligandBcharged.mol2
ligand_outfile_prefix                                flex
limit_max_ligands                                    no
read_mol_solvation                                   no
write_orientations                                   no
write_conformations                                  no
skip_molecule                                        no
calculate_rmsd                                       yes
use_rmsd_reference_mol                               1
rank_ligands                                         yes
max_ranked_ligands                                   3
scored_conformer_output_override                     no
Orient Ligand Parameters
-----------------------------------------------------------------------
use_rmsd_reference_mol                               1
rank_ligands                                         yes
max_ranked_ligands                                   3
scored_conformer_output_override                     no
```

```
Orient Ligand Parameters
-------------------------------------------------------------------
orient_ligand                                                  yes
automated_matching                                             yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                               600
critical_points                                                no
chemical_matching                                              no
use_ligand_spheres                                             no

Flexible Ligand Parameters
-------------------------------------------------------------------
flexible_ligand                                                yes
min_anchor_size                                                40
num_anchor_orients_for_growth                                  100
number_confs_for_next_growth                                   100
use_internal_energy                                            yes
internal_energy_att_exp                                        6
internal_energy_rep_exp                                        12
internal_energy_dielectric                                     4.0
use_clash_overlap                                              no

Bump Filter Parameters
-------------------------------------------------------------------
bump_filter                                                    no

Master Score Parameters
-------------------------------------------------------------------
score_molecules                                                yes
Contact Score Paramters
-------------------------------------------------------------------
contact_score_primary                                          no
contact_score_secondary                                        no
Grid Score Parameters
-------------------------------------------------------------------
grid_score_primary                                             yes
grid_score_secondary                                           yes
grid_score_vdw_scale                                           1
grid_score_es_scale                                            1
grid_score_grid_prefix                                         grid
Simplex Minimization Parameters
-------------------------------------------------------------------
minimize_ligand                                                yes
minimize_anchor                                                yes
minimize_flexible_growth                                       yes
use_advanced_simplex_parameters                                no
simplex_max_cycles                                             1
simplex_score_converge                                         0.1
simplex_cycle_converge                                         1.0
simplex_trans_step                                             1.0
simplex_rot_step                                               0.1
simplex_tors_step                                              10.0
simplex_anchor_max_iterations                                  500
simplex_grow_max_iterations                                    500
simplex_final_min_add_internal                                 no
simplex_secondary_minimize_pose                                no
```

```
simplex_random_seed                                              0
Atom Typing Parameters
-----------------------------------------------------------------
atom_model                                                       all
vdw_defn_file
vdw_AMBER_parm99.defn
flex_defn_file                                                   flex.defn
flex_drive_file                                                  flex_drive.tbl
-----------------------------------------------------------------
Initializing Library File Routines...
Initializing Orienting Routines...
Initializing Conformer Generator Routines...
Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg
-----------------------------------
Molecule: ligandB
 Elapsed time:    460 seconds

 Anchors:         1
 Orientations:          600
 Conformations:         89

 Grid Score:           -40.356724
          vdw:           -41.383575
           es:             1.026853

1 Molecules Processed
Total elapsed time:    460 seconds
```

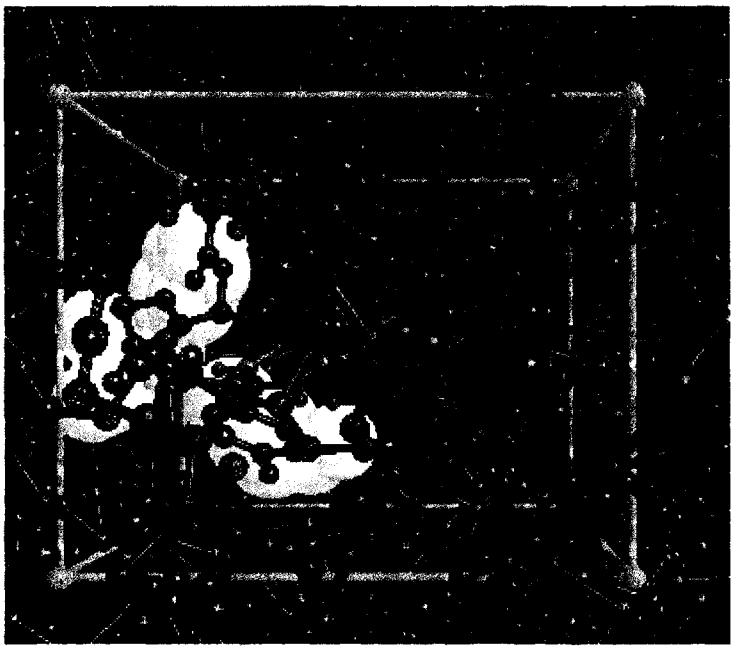


Figure 5. Ligand B in green docked in the red receptor structure COX-2.

File 5. Input file for rigid docking for Ligand B.

```
ligand_atom_file                                    ligandBcharged.mol2
ligand_outfile_prefix                                         rigid
limit_max_ligands                                             no
read_mol_solvation                                            no
write_orientations                                            no
write_conformations                                           no
skip_molecule                                                 no
calculate_rmsd                                                yes
use_rmsd_reference_mol                                        1
rank_ligands                                                  yes
max_ranked_ligands                                            3
scored_conformer_output_override                              no
orient_ligand                                                 yes
automated_matching                                            yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                              1000
critical_points                                               no
chemical_matching                                             no
use_ligand_spheres                                            no
flexible_ligand                                               no
bump_filter                                                   no
score_molecules                                               yes
contact_score_primary                                         no
contact_score_secondary                                       no
grid_score_primary                                            yes
grid_score_secondary                                          yes
grid_score_vdw_scale                                          1
grid_score_es_scale                                           1
grid_score_grid_prefix                                        grid
minimize_ligand                                               yes
simplex_max_iterations                                        1000
simplex_max_cycles                                            1
simplex_score_converge                                        0.1
simplex_cycle_converge                                        1.0
simplex_trans_step                                            1.0
simplex_rot_step                                              0.1
simplex_tors_step                                             10.0
simplex_final_min_add_internal                                no
simplex_secondary_minimize_pose                               no
simplex_random_seed                                           0
atom_model                                                    all
vdw_defn_file                          vdw_AMBER_parm99.defn
flex_defn_file                                      flex.defn
flex_drive_file                      flex_drive.tbl
```

File 6. Output file for Ligand B in rigid docking.

```
DOCK v6.0
Released June 2006
Copyright UCSF
-------------------------------------
Molecule Library Parameters
-----------------------------------------------------------------
ligand_atom_file                          ligandBcharged.mol2
ligand_outfile_prefix                                   rigid
limit_max_ligands                                          no
read_mol_solvation                                         no
write_orientations                                         no
write_conformations                                        no
skip_molecule                                              no
calculate_rmsd                                            yes
use_rmsd_reference_mol                                      1
rank_ligands                                              yes
max_ranked_ligands                                          3
scored_conformer_output_override                           no

Orient Ligand Parameters
-----------------------------------------------------------------
orient_ligand                                             yes
automated_matching                                        yes
receptor_site_file                  5coxselectedspheres.sph
max_orientations                                         1000
critical_points                                            no
chemical_matching                                          no
use_ligand_spheres                                         no

Flexible Ligand Parameters
-----------------------------------------------------------------
flexible_ligand                                            no

Bump Filter Parameters
-----------------------------------------------------------------
bump_filter                                                no

Master Score Parameters
-----------------------------------------------------------------
score_molecules                                           yes

Contact Score Paramters
contact_score_primary                                      no
contact_score_secondary                                    no
```

```
Grid Score Parameters
-----------------------------------------------------------------
grid_score_primary                                          yes
grid_score_secondary                                        yes
grid_score_vdw_scale                                        1
grid_score_es_scale                                         1
grid_score_grid_prefix                                      grid
Simplex Minimization Parameters
-----------------------------------------------------------------
minimize_ligand                                             yes
simplex_max_iterations                                      1000
simplex_max_cycles                                          1
simplex_score_converge                                      0.1
simplex_cycle_converge                                      1.0
simplex_trans_step                                          1.0
simplex_rot_step                                            0.1
simplex_tors_step                                           10.0
simplex_final_min_add_internal                              no
simplex_secondary_minimize_pose                             no
simplex_random_seed                                         0
Atom Typing Parameters
-----------------------------------------------------------------
atom_model                                                  all
vdw_defn_file                          vdw_AMBER_parm99.defn
flex_defn_file                                flex.defn
flex_drive_file                               flex_drive.tbl
-----------------------------------------------------------------
Initializing Library File Routines...
Initializing Orienting Routines...
Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg
-----------------------------------
Molecule: ligandB
Elapsed time:    137 seconds

Anchors:         1
Orientations:          1000
Conformations:         1000

Grid Score:           -35.400383
          vdw:          -36.430904
           es:            1.030523

1 Molecules Processed
Total elapsed time:    138 seconds
```

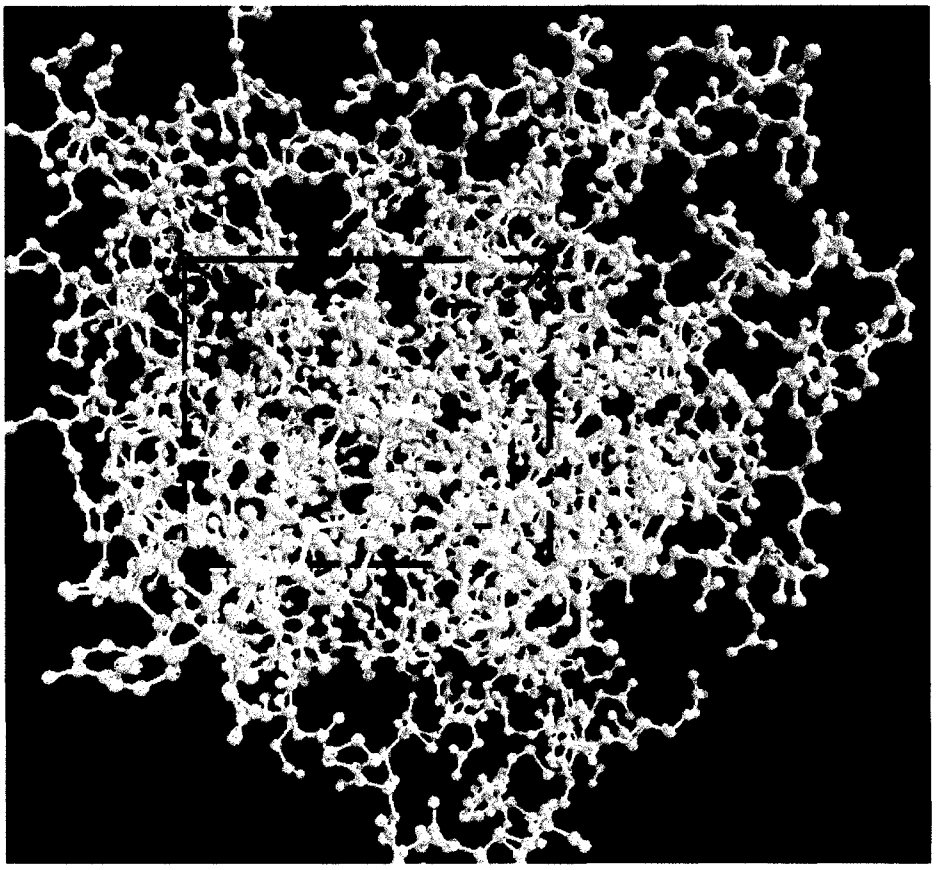


Figure 6. Ligand B docked in white receptor structure COX-2 in rigid docking.

3. Ligand C

The information for Ligand C was obtained from KEGG Ligand database. The mol file was saved in mol2 format and visualized in VEGA. The information and mol file for Ligand C is given in Figure 7 and File 5.

KEGG **DRUG: D00567**

Help

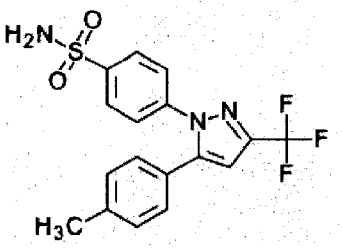
Entry	D00567 Drug
Name	Celecoxib (JAN/USAN/INN); Celebrex (TN)
Formula	C17H14F3N3O2S
Mass	381.0759
Structure	 D00567 Mol file KCF file DB search
Target	cyclooxygenase-2 (COX-2) inhibitor [HSA:5743] [EC:1.14.99.1]
Activity	Anti-inflammatory and analgesic [cyclooxygenase-2 inhibitor]
Remark	Same as: C07589 Therapeutic category: 1149
Pathway	PATH: map07112 1,2-Diphenyl substitution family
Other DBs	CAS: 169590-42-5 PubChem: 7847633
LinkDB	All DBs
KCF data	Show

Figure 7. Information of Ligand C in the KEGG Ligand Database.

File 5. Mol file for Ligand C exported from KEGG Ligand Database.

```
26 28  0  0  0  0  0  0  0  0999 V2000
   23.9641  -17.9834    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   25.3657  -17.9834    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   23.1231  -16.8622    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   21.8150  -17.2826    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   21.8150  -18.6842    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   23.1231  -19.1047    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   25.3657  -19.3850    0.0000 F   0  0  0  0  0  0  0  0  0  0  0  0
   26.7672  -17.9834    0.0000 F   0  0  0  0  0  0  0  0  0  0  0  0
   25.3657  -16.5819    0.0000 F   0  0  0  0  0  0  0  0  0  0  0  0
   18.1709  -15.1803    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   18.1709  -16.5819    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   19.3856  -17.2826    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   20.6003  -16.5819    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   20.6003  -15.1803    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   19.3856  -14.4795    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   18.1709  -19.3850    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   18.1709  -20.7866    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   19.3856  -21.4874    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   20.6003  -20.7866    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   20.6003  -19.3850    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   19.3856  -18.6842    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   16.9572  -21.4876    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   16.9572  -14.4793    0.0000 S   0  0  0  0  0  0  0  0  0  0  0  0
   17.6570  -13.2647    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   16.2554  -15.6942    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   15.7415  -13.7787    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0     0  0
  1  3  2  0     0  0
  3  4  1  0     0  0
  4  5  1  0     0  0
  5  6  2  0     0  0
  1  6  1  0     0  0
  2  7  1  0     0  0
  2  8  1  0     0  0
  2  9  1  0     0  0
 10 15  1  0     0  0
 13  4  1  0     0  0
 16 17  2  0     0  0
 17 18  1  0     0  0
 18 19  2  0     0  0
 19 20  1  0     0  0
 20 21  2  0     0  0
 16 21  1  0     0  0
 20  5  1  0     0  0
 17 22  1  0     0  0
 10 23  1  0     0  0
 23 24  2  0     0  0
 23 25  2  0     0  0
 23 26  1  0     0  0  M  END
```

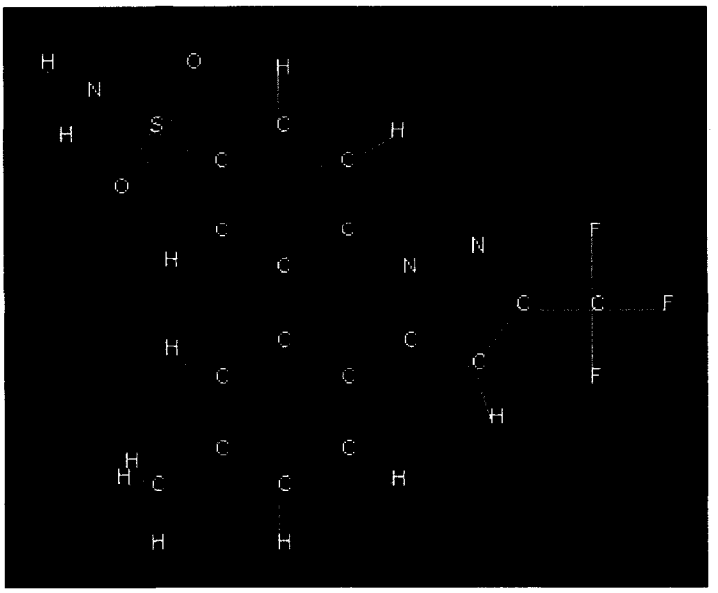


Figure 8. Ligand C with hydrogen atoms added.

File 6. Input file 5coxanchor_and _grow.in for LigandC.

```
ligand_atom_file                                    ligandCcharged.mol2
ligand_outfile_prefix                                                flex
limit_max_ligands                                                    no
read_mol_solvation                                                   no
write_orientations                                                   no
write_conformations                                                  no
skip_molecule                                                        no
calculate_rmsd                                                       yes
use_rmsd_reference_mol                                               1
rank_ligands                                                         no
num_scored_conformers_written                                        1
orient_ligand                                                        yes
automated_matching                                                   yes
receptor_site_file                          5coxselectedspheres.sph
max_orientations                                                     600
critical_points                                                      no
chemical_matching                                                    no
use_ligand_spheres                                                   no
flexible_ligand                                                      yes
min_anchor_size                                      40
num_anchor_orients_for_growth                        100
number_confs_for_next_growth                         100
use_internal_energy                                  yes
internal_energy_att_exp                              6
internal_energy_rep_exp                              12
internal_energy_dielectric                           4.0
use_clash_overlap                                    no
```

```
bump_filter                                                    no
score_molecules                                                yes
contact_score_primary                                          no
contact_score_secondary                                        no
grid_score_primary                                             yes
grid_score_secondary                                           yes
grid_score_vdw_scale                                           1
grid_score_es_scale                                            1
grid_score_grid_prefix                                         grid
minimize_ligand                                                yes
minimize_anchor                                                yes
minimize_flexible_growth                                       yes
use_advanced_simplex_parameters                                no
simplex_max_cycles                                             1
simplex_score_converge                                         0.1
simplex_cycle_converge                                         1.0
simplex_trans_step                                             1.0
simplex_rot_step                                               0.1
simplex_tors_step                                              10.0
simplex_anchor_max_iterations                                  500
simplex_grow_max_iterations                                    500
simplex_final_min_add_internal                                 no
simplex_secondary_minimize_pose                                no
simplex_random_seed                                            0
atom_model                                                     all
vdw_defn_file                          vdw_AMBER_parm99.defn
flex_defn_file                                    flex.defn
flex_drive_file                                flex_drive.tbl
```

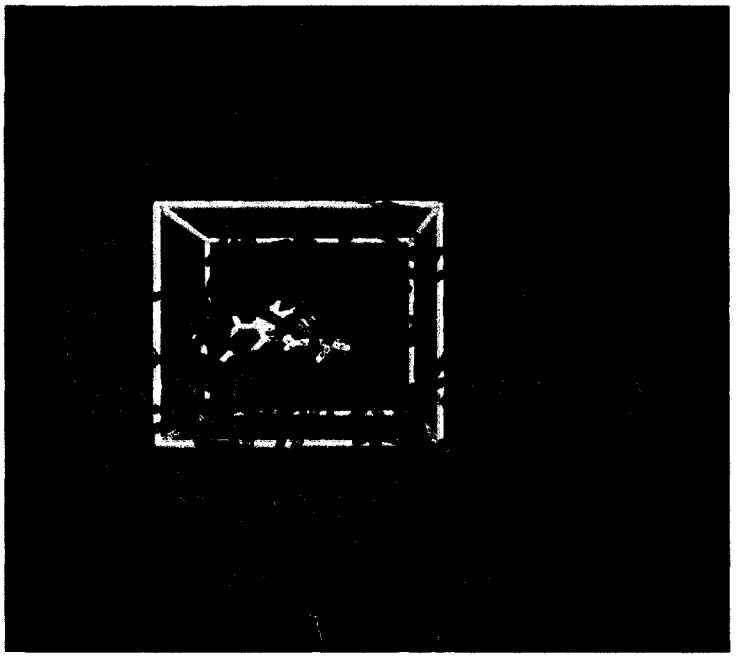


Figure 9. Ligand C in red receptor structure in flexible docking.

File 7. Output file for Ligand C in flexible docking.

```
DOCK v6.0
Released June 2006
Copyright UCSF
-------------------------------------
Molecule Library Parameters
-------------------------------------------------------------------------
ligand_atom_file                                ligandCcharged.mol2
ligand_outfile_prefix                                          flex
limit_max_ligands                                                no
read_mol_solvation                                               no
write_orientations                                               no
write_conformations                                              no
skip_molecule                                    no
calculate_rmsd                                  yes
use_rmsd_reference_mol                           1
rank_ligands                                                     no
num_scored_conformers_written                                    1
Orient Ligand Parameters
-----------------------------------------------------------------
orient_ligand                                                    yes
automated_matching                                               yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                                 600
critical_points                                                  no
chemical_matching                                                no
use_ligand_spheres                                               no

Flexible Ligand Parameters
-----------------------------------------------------------------
flexible_ligand                                                  yes
min_anchor_size                                                  40
num_anchor_orients_for_growth                                    100
number_confs_for_next_growth                                     100
use_internal_energy                                              yes
internal_energy_att_exp                                          6
internal_energy_rep_exp                                          12
internal_energy_dielectric                                       4.0
use_clash_overlap                                                no

Bump Filter Parameters
-----------------------------------------------------------------
bump_filter                                                      no
Master Score Parameters
-----------------------------------------------------------------
score_molecules                                                  yes

Contact Score Paramters
-------------------------------------------------------------------------
contact_score_primary                                            no
contact_score_secondary                                          no
```

```
Grid Score Parameters
-----------------------------------------------------------------------
grid_score_primary                                                  yes
grid_score_secondary                                                yes
grid_score_vdw_scale                                                1
grid_score_es_scale                                                 1
grid_score_grid_prefix                                              grid

Simplex Minimization Parameters
minimize_ligand                                                     yes
minimize_anchor                                                     yes
minimize_flexible_growth                                            yes
use_advanced_simplex_parameters                                     no
simplex_max_cycles                                                  1
simplex_score_converge                                              0.1
simplex_cycle_converge                                              1.0
simplex_trans_step                                                  1.0
simplex_rot_step                                                    0.1
simplex_tors_step                                                   10.0
simplex_anchor_max_iterations                                       500
simplex_grow_max_iterations                                         500
simplex_final_min_add_internal                                      no
simplex_secondary_minimize_pose                                     no
simplex_random_seed                                                 0

Atom Typing Parameters
-----------------------------------------------------------------------
atom_model                                                          all
vdw_defn_file                                       vdw_AMBER_parm99.defn
flex_defn_file                                                  flex.defn
flex_drive_file
flex_drive.tbl
-----------------------------------------------------------------------
Initializing Library File Routines...
Initializing Orienting Routines...
Initializing Conformer Generator Routines...
Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg

---------------------------------
Molecule: ligandC1charged

 Elapsed time:    1523 seconds

 Anchors:         1
 Orientations:          600
 Conformations:         75

 Grid Score:          -34.592548
          vdw:          -25.055355
           es:           -9.537194

1 Molecules Processed
Total elapsed time:    1523 seconds
```

File 8. Input file for Ligand C in rigid docking.

```
ligand_atom_file                          ligandCcharged.mol2
ligand_outfile_prefix                                   rigid
limit_max_ligands                                          no
read_mol_solvation                                         no
write_orientations                                         no
write_conformations                                        no
skip_molecule                                              no
calculate_rmsd                                             no
rank_ligands                                               no
num_scored_conformers_written                              1
orient_ligand                                              yes
automated_matching                                         yes
receptor_site_file                    5coxselectedspheres.sph
max_orientations                                           1000
critical_points                                            no
chemical_matching                                          no
use_ligand_spheres                                         no
flexible_ligand                                            no
bump_filter                                                no
score_molecules                                            yes
contact_score_primary                                      no
contact_score_secondary                                    no
grid_score_primary                                         yes
grid_score_secondary                                       yes
grid_score_vdw_scale                                       1
grid_score_es_scale                                        1
grid_score_grid_prefix                                     grid
minimize_ligand                                            yes
simplex_max_iterations                                     1000
simplex_max_cycles                                         1
simplex_score_converge                                     0.1
simplex_cycle_converge                                     1.0
simplex_trans_step                                         1.0
simplex_rot_step                                           0.1
simplex_tors_step                                          10.0
simplex_final_min_add_internal                             no
simplex_secondary_minimize_pose                            no
simplex_random_seed                                        0
atom_model                                                 all
vdw_defn_file                       vdw_AMBER_parm99.defn
flex_defn_file                            flex.defn
flex_drive_file                        flex_drive.tbl
```

File 9. Output file for Ligand C in rigid docking.

```
DOCK v6.0
Released June 2006
Copyright UCSF
-------------------------------------
Molecule Library Parameters
-----------------------------------------------------------------
ligand_atom_file                        ligandCcharged.mol2
ligand_outfile_prefix                                 rigid
limit_max_ligands                                        no
read_mol_solvation                                       no
write_orientations                                       no
write_conformations                                      no
skip_molecule                                            no
calculate_rmsd                                           no
rank_ligands                                             no
num_scored_conformers_written                            1
Orient Ligand Parameters
-----------------------------------------------------------------
orient_ligand                                            yes
automated_matching                                       yes
receptor_site_file                  5coxselectedspheres.sph
max_orientations                                        1000
critical_points                                          no
chemical_matching                                        no
use_ligand_spheres                                       no

Flexible Ligand Parameters
-----------------------------------------------------------------
flexible_ligand                                          no
Bump Filter Parameters
-----------------------------------------------------------------
bump_filter                                              no
Master Score Parameters
-----------------------------------------------------------------
score_molecules                                          yes
Contact Score Paramters
-----------------------------------------------------------------
contact_score_primary                                    no
contact_score_secondary                                  no

Grid Score Parameters
-----------------------------------------------------------------
grid_score_primary                                       yes
grid_score_secondary                                     yes
grid_score_vdw_scale                                   1
grid_score_es_scale                                      1
grid_score_grid_prefix                                 grid

Simplex Minimization Parameters
-----------------------------------------------------------------
minimize_ligand                                          yes
simplex_max_iterations                                  1000
simplex_max_cycles                                       1
simplex_score_converge                                   0.1
simplex_cycle_converge                                   1.0
simplex_trans_step                                       1.0
```

```
simplex_rot_step                                                    0.1
simplex_tors_step                                                   10.0
simplex_final_min_add_internal                                      no
simplex_secondary_minimize_pose                                     no
simplex_random_seed                                                 0

Atom Typing Parameters
-----------------------------------------------------------------------
atom_model                                                          all
vdw_defn_file
 vdw_AMBER_parm99.defn
flex_defn_file                                                      flex.defn
flex_drive_file                                                flex_drive.tbl
-----------------------------------------------------------------------
Initializing Library File Routines...
Initializing Orienting Routines...
Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg
-----------------------------------
Molecule: ligandC1charged
  Elapsed time:    158 seconds
  Anchors:         1
  Orientations:          1000
  Conformations:         1000
  Grid Score:           -14.554059
            vdw:           -3.643506
             es:           -10.910553

 1 Molecules Processed
 Total elapsed time:    158 seconds
```

Figure 10. Ligand C in the receptor structure (rigid docking).

4. Ligand D

The information for Ligand D was obtained from Protein Data Bank. The ligand that was docked into another structure of cyclooxygenase was used. The file 6COX.pdb was opened in VEGA ZZ program and the ligand SC-558 was isolated and saved as ligand D in mol2 format.

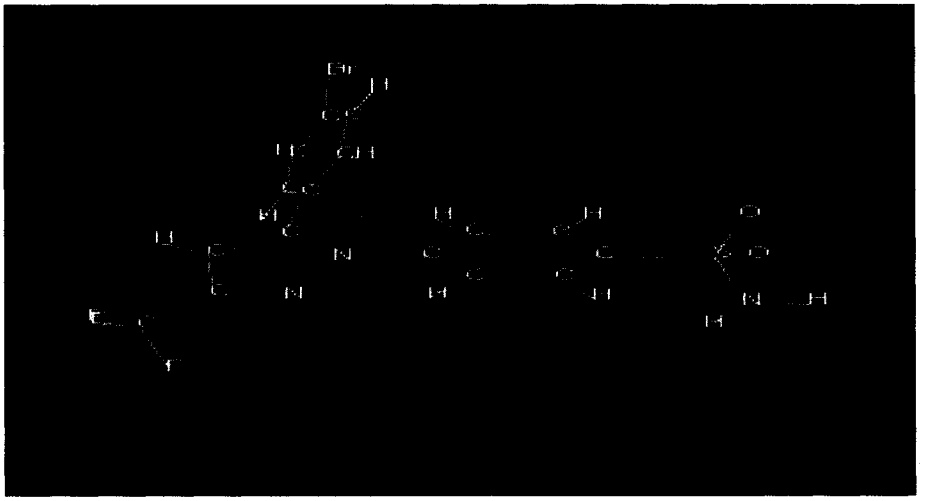


Figure 11. The structure of Ligand D.

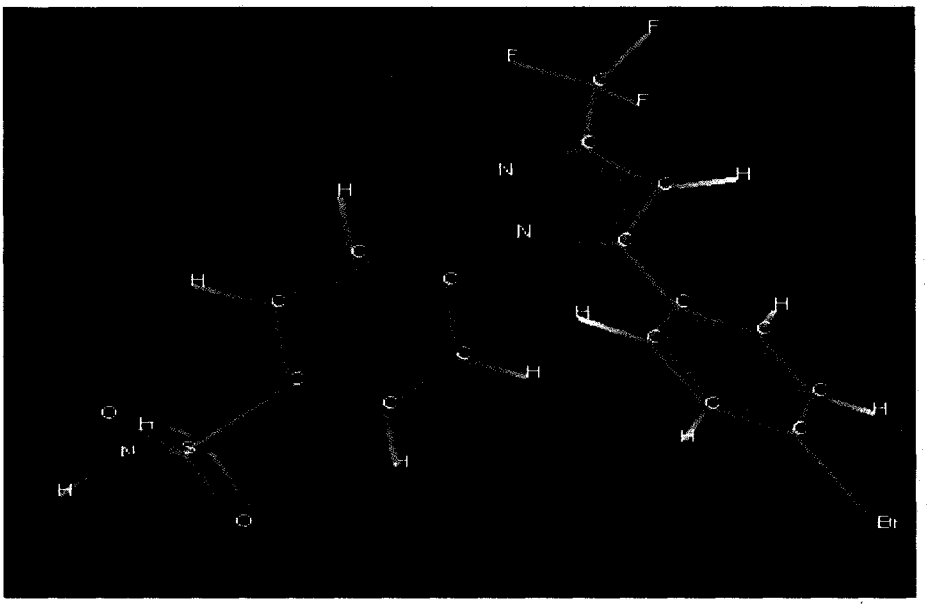


Figure 12. The structure of Ligand D with hydrogens.

File 10. The input file for flexible docking of Ligand D.

```
ligand_atom_file                                    s58charged.mol2
ligand_outfile_prefix                                          flex
limit_max_ligands                                              no
read_mol_solvation                                             no
write_orientations                                             yes
write_conformations                                            yes
skip_molecule                                                  no
calculate_rmsd                                                 yes
use_rmsd_reference_mol                              s58charged.mol2
rank_ligands                                                   yes
max_ranked_ligands                                             20
scored_conformer_output_override                               no
orient_ligand                                                  yes
automated_matching                                             yes
receptor_site_file                          5coxselectedspheres.sph
max_orientations                                               100
critical_points                                                no
chemical_matching                                              no
use_ligand_spheres                                             no
flexible_ligand                                                yes
min_anchor_size                                                40
num_anchor_orients_for_growth                                  100
number_confs_for_next_growth                                   100
use_internal_energy                                            yes
internal_energy_att_exp                                        6
internal_energy_rep_exp                                        12
internal_energy_dielectric                                     4.0
use_clash_overlap                                              no
bump_filter                                                    no
score_molecules                                                yes
contact_score_primary                                          no
contact_score_secondary                                        no
grid_score_primary                                             yes
grid_score_secondary                                           yes
grid_score_vdw_scale                                           1
grid_score_es_scale                                            1
grid_score_grid_prefix                                         grid
minimize_ligand                                                yes
minimize_anchor                                                yes
minimize_flexible_growth                                       yes
use_advanced_simplex_parameters                                no
simplex_max_cycles                                             1
simplex_score_converge                                         0.1
simplex_cycle_converge                                         1.0
simplex_trans_step                                             1.0
simplex_rot_step                                               0.1
simplex_tors_step                                              10.0
simplex_anchor_max_iterations                                  500
```

```
simplex_grow_max_iterations                                    500
simplex_final_min_add_internal                                 no
simplex_secondary_minimize_pose                                no
simplex_random_seed                                            0
atom_model                                                     all
vdw_defn_file                          vdw_AMBER_parm99.defn
flex_defn_file                                flex.defn
flex_drive_file                            flex_drive.tbl
```

File 11.The output file of the flexible docked Ligand D.

```
DOCK v6.0
Released June 2006
Copyright UCSF
-------------------------------------
Molecule Library Parameters
-----------------------------------------------------------------------
ligand_atom_file
s58charged.mol2
ligand_outfile_prefix                                          flex
limit_max_ligands                                              no
read_mol_solvation                                             no
write_orientations                                             no
write_conformations                                            no
skip_molecule                                                  no
calculate_rmsd                                                 no
rank_ligands                                                   no
num_scored_conformers_written                                  1

Orient Ligand Parameters
-----------------------------------------------------------------------
orient_ligand                                                  yes
automated_matching                                             yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                               500
critical_points                                                no
chemical_matching                                              no
use_ligand_spheres                                             no

Flexible Ligand Parameters
-----------------------------------------------------------------------
flexible_ligand                                                yes
min_anchor_size                                                40
num_anchor_orients_for_growth                                  100
number_confs_for_next_growth                                   100
use_internal_energy                                            yes
internal_energy_att_exp                                        6
internal_energy_rep_exp                                        12
internal_energy_dielectric                                     4.0
use_clash_overlap                                              no
```

```
Bump Filter Parameters
-----------------------------------------------------------------------
bump_filter                                                        no
Master Score Parameters
-----------------------------------------------------------------------
score_molecules                                                    yes
Contact Score Paramters
-----------------------------------------------------------------------
contact_score_primary                                              no
contact_score_secondary                                            no
Grid Score Parameters
-----------------------------------------------------------------------
grid_score_primary                                                 yes
grid_score_secondary                                               yes
grid_score_vdw_scale                                               1
grid_score_es_scale                                                1
grid_score_grid_prefix                                             grid

Simplex Minimization Parameters
-----------------------------------------------------------------------
minimize_ligand                                                    yes
minimize_anchor                                                    yes
minimize_flexible_growth                                           yes
use_advanced_simplex_parameters                                    no
simplex_max_cycles                                                 1
simplex_score_converge                                             0.1
simplex_cycle_converge                                             1.0
simplex_trans_step                                                 1.0
simplex_rot_step                                                   0.1
simplex_tors_step                                                  10.0
simplex_anchor_max_iterations                                      500
simplex_grow_max_iterations                                        500
simplex_final_min_add_internal                                     no
simplex_secondary_minimize_pose                                    no
simplex_random_seed                                                0
Atom Typing Parameters atom_model                                 all
vdw_defn_file                                vdw_AMBER_parm99.defn
flex_defn_file                                       flex.defn
flex_drive_file
flex_drive.tbl
-----------------------------------------------------------------------
Initializing Library File Routines...
Initializing Orienting Routines...
Initializing Conformer Generator Routines...
Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg
---------------------------------
Molecule: s58charged
 Elapsed time:     494 seconds

 Anchors:          1
 Orientations:           600
 Conformations:          46

 Grid Score:           -40.702839
          vdw:           -27.807253
           es:           -12.895587
1 Molecules Processed
Total elapsed time:      494 seconds
```

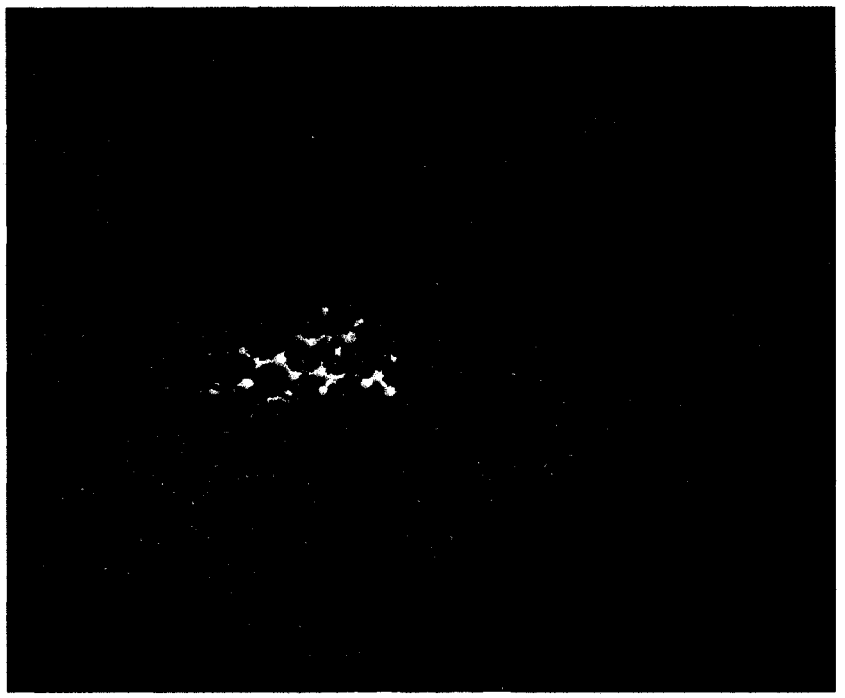


Figure 13. Ligand D bound in red receptor structure in flexible docking.

Figure 14. Comparison of 2 model structures COX-2(pink) from the demo module of UCSF Chimera vs. the model determined here for Ligand D SC-558(grey).

File 12. The input file for rigid docking of Ligand D.

```
ligand_atom_file                                    s58charged.mol2
ligand_outfile_prefix                                         rigid
limit_max_ligands                                             no
read_mol_solvation                                            no
write_orientations                                            no
write_conformations                                           no
skip_molecule                                                 no
calculate_rmsd                                                no
rank_ligands                                                  no
num_scored_conformers_written                                 1
orient_ligand                                                 yes
automated_matching                                            yes
receptor_site_file                          5coxselectedspheres.sph
max_orientations                                              1000
critical_points                                               no
chemical_matching                                             no
use_ligand_spheres                                            no
flexible_ligand                                               no
bump_filter                                                   no
score_molecules                                               yes
contact_score_primary                                         no
contact_score_secondary                                       no
grid_score_primary                                            yes
grid_score_secondary                                          yes
grid_score_vdw_scale                                          1
grid_score_es_scale                                           1
grid_score_grid_prefix                                        grid
minimize_ligand                                               yes
simplex_max_iterations                                        1000
simplex_max_cycles                                            1
simplex_score_converge                                        0.1
simplex_cycle_converge                                        1.0
simplex_trans_step                                            1.0
simplex_rot_step                                              0.1
simplex_tors_step                                             10.0
simplex_final_min_add_internal                                no
simplex_secondary_minimize_pose                               no
simplex_random_seed                                           0
atom_model                                                    all
vdw_defn_file                          vdw_AMBER_parm99.defn
flex_defn_file                                       flex.defn
flex_drive_file
flex_drive.tbl
```

File 13. The output file for rigid docking of Ligand D.

```
---------------------------------------
DOCK v6.0

Released June 2006
Copyright UCSF
---------------------------------------
Molecule Library Parameters
------------------------------------------------------------------
ligand_atom_file
s58charged.mol2
ligand_outfile_prefix                                          rigid
limit_max_ligands                                              no
read_mol_solvation                                             no
write_orientations                                             no
write_conformations                                            no
skip_molecule                                                  no
calculate_rmsd                                                 no
rank_ligands                                                   no
num_scored_conformers_written                                  1

Orient Ligand Parameters
------------------------------------------------------------------
orient_ligand                                                  yes
automated_matching                                             yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                               1000
critical_points                                                no
chemical_matching                                              no
use_ligand_spheres                                             no

Flexible Ligand Parameters
------------------------------------------------------------------
flexible_ligand                                                no

Bump Filter Parameters
------------------------------------------------------------------
bump_filter                                                    no

Master Score Parameters
------------------------------------------------------------------
score_molecules                                                yes
Contact Score Paramters
------------------------------------------------------------------
contact_score_primary                                          no
contact_score_secondary                                        no
```

```
Grid Score Parameters
grid_score_primary                                              yes
grid_score_secondary                                            yes
grid_score_vdw_scale                                            1
grid_score_es_scale                                             1
grid_score_grid_prefix                                          grid

Simplex Minimization Parameters
-----------------------------------------------------------------
minimize_ligand                                                 yes
simplex_max_iterations                                          1000
simplex_max_cycles                                              1
simplex_score_converge                                          0.1
simplex_cycle_converge                                          1.0
simplex_trans_step                                              1.0
simplex_rot_step                                                0.1
simplex_tors_step                                               10.0
simplex_final_min_add_internal                                  no
simplex_secondary_minimize_pose                                 no
simplex_random_seed                                             0

Atom Typing Parameters
-----------------------------------------------------------------
atom_model
all
vdw_defn_file
vdw_AMBER_parm99.defn
flex_defn_file
flex.defn
flex_drive_file
flex_drive.tbl
-----------------------------------------------------------------
Initializing Library File Routines...
Initializing Orienting Routines...
Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg
-----------------------------------
Molecule: s58charged

 Elapsed time:    89 seconds

 Anchors:         1
 Orientations:          1000
 Conformations:         1000

 Grid Score:           13.429971
          vdw:          26.112669
           es:          -12.682698

1 Molecules Processed
Total elapsed time:     89 seconds
```

Figure 15. Ligand D docked in red receptor structure COX-2 in rigid docking.

5. Ligand E

The information for Ligand Ewas obtained from KEGG Ligand database. The mol file was saved in mol2 format and visualized in VEGA. The information and mol file for Ligand E is given in Figure 16 and File 5.

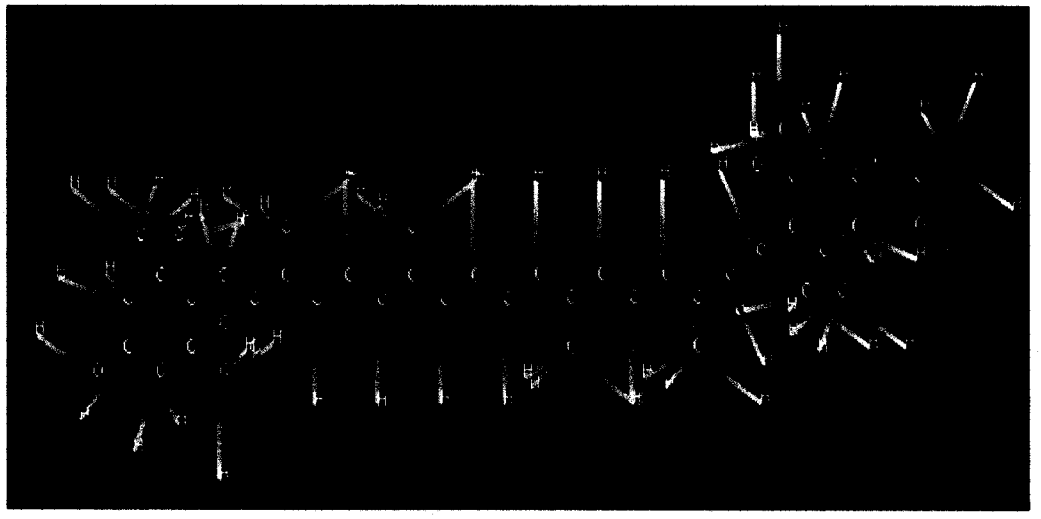


Figure 16. Ligand E structure with hydrogens.

KEGG COMPOUND: C08596

Help

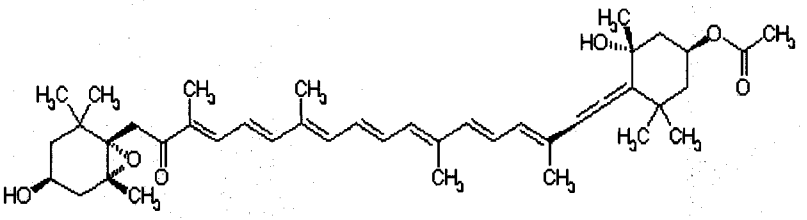
Entry	C08596 Compound
Name	Fucoxanthin
Formula	C42H58O6
Mass	658.4233
Structure	 C08596 Mol file KCF file DB search Jmol KegDraw
Other DBs	CAS: 3351-86-8 PubChem: 10789 ChEBI: 5186
LinkDB	All DBs
KCF data	Show

Figure 17. Information for Ligand E in the KEGG Ligand Database.

File 14. The mol file for Ligand E.

```
ISISHOST03240423202D 1   1.00000     0.00000  7636

 48 50  0     1  0            999 V2000
   -4.6414   -0.3483    0.0000 C   0  0  1  0  0  0           0  0  0
   -4.6414   -0.8759    0.0000 C   0  0  2  0  0  0           0  0  0
   -4.1931   -0.6103    0.0000 O   0  0  0  0  0  0           0  0  0
   -5.1000   -0.0828    0.0000 C   0  0  3  0  0  0           0  0  0
   -4.1931   -0.0828    0.0000 C   0  0  0  0  0  0           0  0  0
   -5.1000   -1.1379    0.0000 C   0  0  0  0  0  0           0  0  0
   -4.1931   -1.1379    0.0000 C   0  0  0  0  0  0           0  0  0
   -5.5655   -0.3483    0.0000 C   0  0  0  0  0  0           0  0  0
   -5.3690    0.3724    0.0000 C   0  0  0  0  0  0           0  0  0
   -4.8414    0.3724    0.0000 C   0  0  0  0  0  0           0  0  0
   -3.7345   -0.3448    0.0000 C   0  0  0  0  0  0           0  0  0
   -5.5655   -0.8759    0.0000 C   0  0  2  0  0  0           0  0  0
   -3.2828   -0.0793    0.0000 C   0  0  0  0  0  0           0  0  0
   -3.7414   -0.8690    0.0000 O   0  0  0  0  0  0           0  0  0
   -6.0000   -1.1345    0.0000 O   0  0  0  0  0  0           0  0  0
   -2.8276   -0.3414    0.0000 C   0  0  0  0  0  0           0  0  0
   -3.2897    0.4483    0.0000 C   0  0  0  0  0  0           0  0  0
   -2.3724   -0.0759    0.0000 C   0  0  0  0  0  0           0  0  0
   -1.9172   -0.3379    0.0000 C   0  0  0  0  0  0           0  0  0
   -1.4621   -0.0724    0.0000 C   0  0  0  0  0  0           0  0  0
   -1.0069   -0.3345    0.0000 C   0  0  0  0  0  0           0  0  0
   -1.4655    0.4517    0.0000 C   0  0  0  0  0  0           0  0  0
   -0.5517   -0.0724    0.0000 C   0  0  0  0  0  0           0  0  0
   -0.0966   -0.3345    0.0000 C   0  0  0  0  0  0           0  0  0
    0.3552   -0.0690    0.0000 C   0  0  0  0  0  0           0  0  0
    0.8103   -0.3310    0.0000 C   0  0  0  0  0  0           0  0  0
    1.2621   -0.0655    0.0000 C   0  0  0  0  0  0           0  0  0
    0.8069   -0.8586    0.0000 C   0  0  0  0  0  0           0  0  0
    1.7207   -0.3276    0.0000 C   0  0  0  0  0  0           0  0  0
    2.1724   -0.0586    0.0000 C   0  0  0  0  0  0           0  0  0
    2.6276   -0.3241    0.0000 C   0  0  0  0  0  0           0  0  0
    3.0862   -0.0586    0.0000 C   0  0  0  0  0  0           0  0  0
    2.6241   -0.8517    0.0000 C   0  0  0  0  0  0           0  0  0
    3.5379    0.2069    0.0000 C   0  0  0  0  0  0           0  0  0
    3.9897    0.4724    0.0000 C   0  0  0  0  0  0           0  0  0
    4.4448    0.2103    0.0000 C   0  0  3  0  0  0           0  0  0
    3.9897    1.0000    0.0000 C   0  0  2  0  0  0           0  0  0
    4.9034    0.4724    0.0000 C   0  0  0  0  0  0           0  0  0
    4.7069   -0.2448    0.0000 C   0  0  0  0  0  0           0  0  0
    4.1793   -0.2448    0.0000 C   0  0  0  0  0  0           0  0  0
    4.4448    1.2621    0.0000 C   0  0  0  0  0  0           0  0  0
    3.8517    1.5103    0.0000 C   0  0  0  0  0  0           0  0  0
    3.4793    1.1379    0.0000 O   0  0  0  0  0  0           0  0  0
    4.9034    1.0000    0.0000 C   0  0  1  0  0  0           0  0  0
    5.3621    1.2655    0.0000 O   0  0  0  0  0  0           0  0  0
    5.8138    1.0035    0.0000 C   0  0  0  0  0  0           0  0  0
    6.2655    1.2655    0.0000 C   0  0  0  0  0  0           0  0  0
    5.8069    0.4759    0.0000 O   0  0  0  0  0  0           0  0  0
```

```
1  2  1  0     0  0
  1  3  1  6     0  0
  1  4  1  0     0  0
  1  5  1  1     0  0
  2  6  1  0     0  0
  2  7  1  1     0  0
  4  8  1  0     0  0
4  9  1  0     0  0
  4 10  1  0     0  0
  5 11  1  0     0  0
  6 12  1  0     0  0
 11 13  1  0     0  0
 11 14  2  0     0  0
 12 15  1  1     0  0
 13 16  2  0     0  0
 13 17  1  0     0  0
 16 18  1  0     0  0
 18 19  2  0     0  0
 19 20  1  0     0  0
 20 21  2  0     0  0
 20 22  1  0     0  0
 21 23  1  0     0  0
 23 24  2  0     0  0
 24 25  1  0     0  0
 25 26  2  0     0  0
 26 27  1  0     0  0
 26 28  1  0     0  0
 27 29  2  0     0  0
 29 30  1  0     0  0
 30 31  2  0     0  0
 31 32  1  0     0  0
 31 33  1  0     0  0
 32 34  2  0     0  0
 34 35  2  0     0  0
 35 36  1  0     0  0
 35 37  1  0     0  0
 36 38  1  0     0  0
 36 39  1  0     0  0
 36 40  1  0     0  0
 37 41  1  0     0  0
 37 42  1  1     0  0
 37 43  1  6     0  0
 38 44  1  0     0  0
 44 45  1  1     0  0
 45 46  1  0     0  0
 46 47  1  0     0  0
 46 48  2  0     0  0
  2  3  1  6     0  0
  8 12  1  0     0  0
 41 44  1  0     0  0
M  END
```

File 15. The input file for flexible docking of Ligand E.

```
ligand_atom_file                                fucoxanthin.mol2
ligand_outfile_prefix                                      flex
limit_max_ligands                                          no
read_mol_solvation                                         no
write_orientations                                         no
write_conformations                                        no
skip_molecule                                              no
calculate_rmsd                                             no
rank_ligands                                               no
num_scored_conformers_written                              1
orient_ligand                                              yes
automated_matching                                         yes
receptor_site_file                  5coxselectedspheres.sph
max_orientations                                           500
critical_points                                            no
chemical_matching                                          no
use_ligand_spheres                                         no
flexible_ligand                                            yes
min_anchor_size                                            40
num_anchor_orients_for_growth                              100
number_confs_for_next_growth                               100
use_internal_energy                                        yes
internal_energy_att_exp                                    6
internal_energy_rep_exp                                    12
internal_energy_dielectric                                 4.0
use_clash_overlap                                          no
bump_filter                                                no
score_molecules                                            yes
contact_score_primary                                      no
contact_score_secondary                                    no
grid_score_primary                                         yes
grid_score_secondary                                       yes
grid_score_vdw_scale                                       1
grid_score_es_scale                                        1
grid_score_grid_prefix                                     grid
minimize_ligand                                            yes
minimize_anchor                                            yes
minimize_flexible_growth                                   yes
use_advanced_simplex_parameters                            no
simplex_max_cycles                                         1
simplex_score_converge                                     0.1
simplex_cycle_converge                                     1.0
simplex_trans_step                                         1.0
simplex_rot_step                                           0.1
simplex_tors_step                                          10.0
```

```
simplex_anchor_max_iterations                                    500
simplex_grow_max_iterations                                      500
simplex_final_min_add_internal                                   no
simplex_secondary_minimize_pose                                  no
simplex_random_seed                                              0
atom_model                                                       all
vdw_defn_file                                vdw_AMBER_parm99.defn
flex_defn_file                                             flex.defn
flex_drive_file                                      flex_drive.tbl
```

File 15. The output file for flexible docking Ligand E.

```
----------------------------------------
DOCK v6.0

Released June 2006
Copyright UCSF
----------------------------------------
Molecule Library Parameters
----------------------------------------------------------------------
ligand_atom_file
fucoxanthin.mol2
ligand_outfile_prefix                                            flex
limit_max_ligands                                                no
read_mol_solvation                                               no
write_orientations                                               no
write_conformations                                              no
skip_molecule                                                    no
calculate_rmsd                                                   no
rank_ligands                                                     no
num_scored_conformers_written                                    1

Orient Ligand Parameters
----------------------------------------------------------------------
orient_ligand                                                    yes
automated_matching                                               yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                                 500
critical_points                                                  no
chemical_matching                                                no
use_ligand_spheres                                               no
```

```
Flexible Ligand Parameters
-----------------------------------------------------------------
flexible_ligand                                                yes
min_anchor_size                                                40
num_anchor_orients_for_growth                                  100
number_confs_for_next_growth                                   100
use_internal_energy                                            yes
internal_energy_att_exp                                        6
internal_energy_rep_exp                                        12
internal_energy_dielectric                                     4.0
use_clash_overlap                                              no

Bump Filter Parameters
-----------------------------------------------------------------
bump_filter                                                    no

Master Score Parameters
-----------------------------------------------------------------
score_molecules                                                yes

Contact Score Paramters
-----------------------------------------------------------------
contact_score_primary                                          no
contact_score_secondary                                        no

Grid Score Parameters
-----------------------------------------------------------------
grid_score_primary                                             yes
grid_score_secondary                                           yes
grid_score_vdw_scale                                           1
grid_score_es_scale                                            1
grid_score_grid_prefix                                         grid

Simplex Minimization Parameters
-----------------------------------------------------------------
minimize_ligand                                                yes
minimize_anchor                                                yes
minimize_flexible_growth                                       yes
use_advanced_simplex_parameters                                no
simplex_max_cycles                                             1
simplex_score_converge                                         0.1
simplex_cycle_converge                                         1.0
simplex_trans_step                                             1.0
simplex_rot_step                                               0.1
simplex_tors_step                                              10.0
simplex_anchor_max_iterations                                  500
```

```
simplex_grow_max_iterations                                    500
simplex_final_min_add_internal                                 no
simplex_secondary_minimize_pose                                no
simplex_random_seed                                            0
Atom Typing Parameters
atom_model                                                     all
vdw_defn_file
vdw_AMBER_parm99.defn
flex_defn_file
flex.defn
flex_drive_file
flex_drive.tbl
----------------------------------------------------------------

Initializing Library File Routines...
Initializing Orienting Routines...
Initializing Conformer Generator Routines...
Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg

-----------------------------------
Molecule: fucoxanthin

 Elapsed time:   2 seconds

 ERROR:  Could not complete growth.
         Confirm grid box is large enough to contain
         ligand and try increasing max_orients.


1 Molecules Processed
Total elapsed time:    2 seconds
```

File16. Input file for rigid docking of Ligand E.

```
ligand_atom_file
fucoxanthin.mol2
ligand_outfile_prefix                                    rigid
limit_max_ligands                                        no
read_mol_solvation                                       no
write_orientations                                       no
write_conformations                                      no
skip_molecule                                            no
calculate_rmsd                                           no
rank_ligands                                             no
num_scored_conformers_written                            1
orient_ligand                                            yes
automated_matching                                       yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                         1000
critical_points                                          no
chemical_matching                                        no
use_ligand_spheres                                       no
flexible_ligand                                          no
bump_filter                                              no
score_molecules                                          yes
contact_score_primary                                    no
contact_score_secondary                                  no
grid_score_primary                                       yes
grid_score_secondary                                     yes
grid_score_vdw_scale                                     1
grid_score_es_scale                                      1
grid_score_grid_prefix                                   grid
minimize_ligand                                          yes
simplex_max_iterations                                   1000
simplex_max_cycles                                       1
simplex_score_converge                                   0.1
simplex_cycle_converge                                   1.0
simplex_trans_step                                       1.0
simplex_rot_step                                         0.1
simplex_tors_step                                        10.0
simplex_final_min_add_internal                           no
simplex_secondary_minimize_pose                          no
atom_model                                               all
vdw_defn_file                          vdw_AMBER_parm99.defn
flex_defn_file                                   flex.defn
flex_drive_file                        flex_drive.tbl
```

File16. Output file for rigid docking of Ligand E

```
---------------------------------------
DOCK v6.0

Released June 2006
Copyright UCSF
---------------------------------------
Molecule Library Parameters
----------------------------------------------------------------------
ligand_atom_file
fucoxanthin.mol2
ligand_outfile_prefix                                          rigid
limit_max_ligands                                              no
read_mol_solvation                                             no
write_orientations                                             no
write_conformations                                            no
skip_molecule                                                  no
calculate_rmsd                                                 no
rank_ligands                                                   no
num_scored_conformers_written                                  1

Orient Ligand Parameters
----------------------------------------------------------------------
orient_ligand                                                  yes
automated_matching                                             yes
receptor_site_file
5coxselectedspheres.sph
max_orientations                                               1000
critical_points                                                no
chemical_matching                                              no
use_ligand_spheres                                             no

Flexible Ligand Parameters
----------------------------------------------------------------------
flexible_ligand                                                no

Bump Filter Parameters
----------------------------------------------------------------------
bump_filter                                                    no
Master Score Parameters
----------------------------------------------------------------------
score_molecules                                                yes

Contact Score Paramters
----------------------------------------------------------------------
contact_score_primary                                          no
contact_score_secondary                                        no
```

```
Grid Score Parameters
-----------------------------------------------------------------------
grid_score_primary                                                  yes
grid_score_secondary                                                yes
grid_score_vdw_scale                                                1
grid_score_es_scale                                                 1
grid_score_grid_prefix                                              grid

Simplex Minimization Parameters
-----------------------------------------------------------------------
minimize_ligand                                                     yes
simplex_max_iterations                                              1000
simplex_max_cycles                                                  1
simplex_score_converge                                              0.1
simplex_cycle_converge                                              1.0
simplex_trans_step                                                  1.0
simplex_rot_step                                                    0.1
simplex_tors_step                                                   10.0
simplex_final_min_add_internal                                      no
simplex_secondary_minimize_pose                                     no
simplex_random_seed                                                 0

Atom Typing Parameters
-----------------------------------------------------------------------
atom_model                                                          all
vdw_defn_file
vdw_AMBER_parm99.defn
flex_defn_file
flex.defn
flex_drive_file
flex_drive.tbl
-----------------------------------------------------------------------

Initializing Library File Routines...
Initializing Orienting Routines...
Initializing Grid Score Routines...
 Reading the energy grid from grid.nrg
-----------------------------------
Molecule: fucoxanthin

 Elapsed time:   7 seconds

 Anchors:        1
 Orientations:         1000
 Conformations:        1000

ERROR:  Conformation could not be scored by DOCK.
Conformation not completely within grid box.
1 Molecules Processed
Total elapsed time:    7 seconds
```

5. DMS Program

To generate spheres in the receptor structure COX-2, molecular surface is generated using the Dot Molecular Surface program (DMS). The input file for DMS is 5coxrec_noH.pdb. The procedure for DMS was followed from the "Tutorial Generating Spheres" website which is a part of the official DOCK website; (http://dock.compbio.ucsf.edu/DOCK_6/tutorials/sphere_generation/generating_spheres.htm). To learn more about the DMS program one can access the website http://www.cgl.ucsf.edu/chimera/docs/UsersGuide/midas/dms1.html.

APPENDIX B

STRUCTURE-BASED DRUG DESIGN LAB MODULE

DRUG TARGET IDENTIFICATION MODULE-1

This module aims to identify the right drug target based on given information. It consists of details explaining the disease process and the drug targets associated with it along with their function. Upon reading the information given, there are questions at the end of the module for students to answer and then determine the best drug target.

1.1 Background

The structure-based drug design process involves detailed knowledge of binding sites of targets (such as proteins) associated with the disease. Every drug works based on structural interaction with the receptor or target molecule. The most common model is the one in which the drug molecule fits itself into the crevice of the target protein like a key in a lock. This strategy results in inhibition of the protein's function, and ultimately halts the progress of the disease.

The first step in the structure-based drug design process starts with drug target identification. A drug target is a macromolecule that is essential for a disease causing agent to thrive. Every such agent has molecular components such as enzymes, and receptors that allow it to multiply or move to different parts of the body. To treat a disease, a specific macromolecule needs to be inactivated. In order to inactivate the drug target, the macromolecule or receptor needs to be identified based on how critical the target is to the disease process. The target should be located in a critical step of the metabolic pathway and its inhibition should consequently kill the disease causing

organism. Drug targets should be unique and specific so that no other pathway is suppressed.

The disease state taken up for study in this module is Osteoarthritis. Osteoarthritis is a chronic disease of the joints involving breakdown of the joint tissue, primarily the cartilage. Cartilage is a tough elastic tissue covering the bone that acts as a shock absorber and keeps the bones from rubbing against each other. The entire joint is enclosed in a capsule that is lined by an inner membrane known as the synovial membrane. The membrane fills the space around the joints with a fluid called the synovial fluid. The synovial fluid nourishes the cartilage and keeps the joints lubricated making movement smooth and easy. Enclosing the joints are muscles, tendons and ligaments. These structures provide support and assist movement in the correct direction. Figure 1 shows the difference between a normal joint and a degenerative joint.

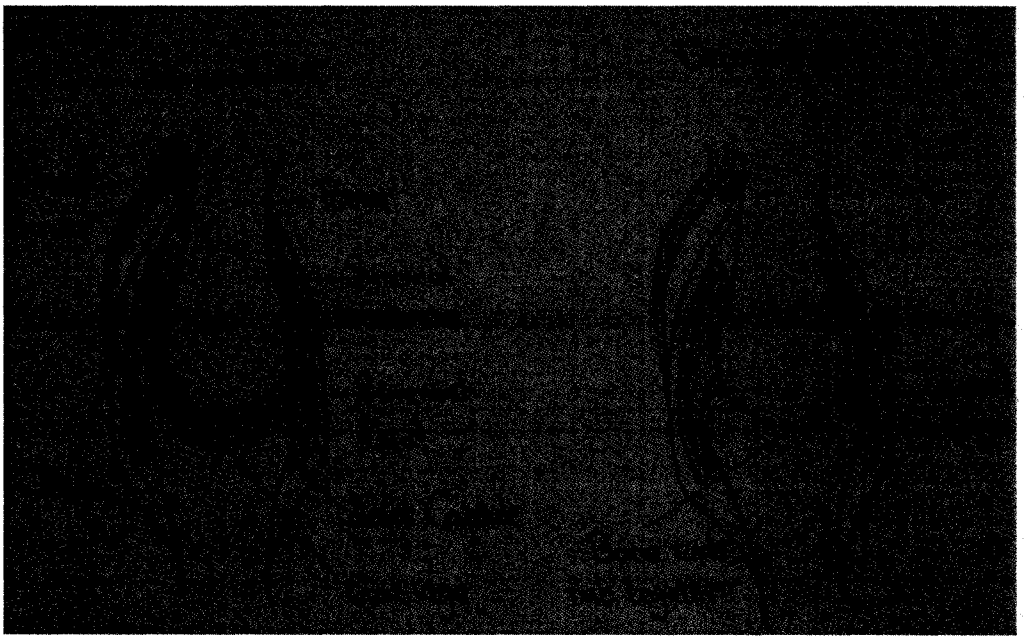


Figure 1. Comparison of a normal joint and degenerative joints[diagram by author].

The degeneration of the cartilage causes inflammation. Inflammation is one of the ways in which the body protects and repairs itself. The inflammatory reaction begins with tissue irritation. In response to the tissue irritation, white blood cells rush from the bloodstream to the area and heal the injury. They release enzymes and active products that affect the nearby cells and alter the blood circulation. The active, toxic products are prostaglandins and leukotrines. They are the potent producers of inflammation. They increase the flow of fluids and white blood cells around the area of inflammation and release more prostaglandins thereby feeding the process. The accumulation of these fluids and cells causes swelling along with pain.

Treatment for osteoarthritis is aimed at controlling pain, improving and maintaining movement, and preventing degeneration of the joints. Pain control is a vital part of the treatment as the drugs help the patient gain control over the effects of the disease. Possible drug targets that were chosen are as follows.

1. Cyclooxygenase enzyme-1 (COX-1)
2. Cyclooxygenase enzyme-2 (COX-2)
3. Prostaglandins
4. Interleukin-1 receptor

1.2 Drug Targets and their Contribution to Osteoarthritis

Prostaglandins are hormones that act as chemical messengers inside the cell activating inflammatory responses and strengthening pain signals. They are unsaturated carboxylic acids consisting of a twenty-carbon skeleton and a five-member ring. They

are synthesized from the fatty acid arachidonate and initial reaction in their production is catalyzed by cyclooxygenase.

There are two different cyclooxygenases, COX-1 and COX-2. Both of the enzymes perform the first step in production of prostaglandins by adding two oxygen molecules to arachidonate. The COX-1 and COX-2 enzymes are similar in structure and have the same substrate (arachidonate). The kinetics of both isoforms is similar but they obtain arachidonate from different sources. Figure 2 shows COX-1 (cyan) and COX-2 (pink) structures viewed in the UCSF Chimera program. Figure 3 shows the COX-1 and COX-2 structures superimposed to illustrate that they are very similar.

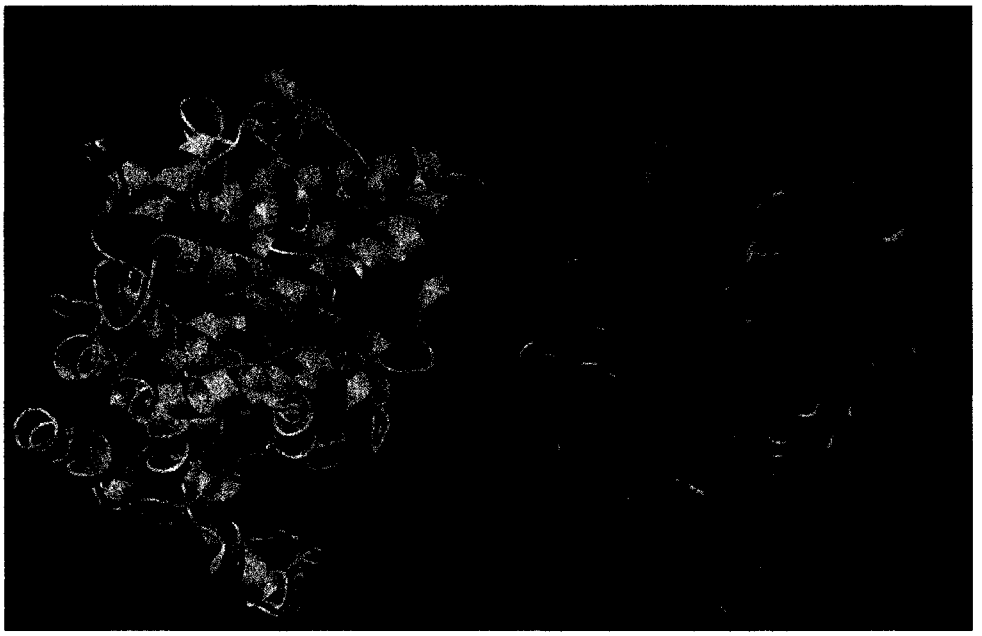


Figure 2. COX-1 and COX-2 structures viewed using UCSF Chimera.

Figure 3. COX-1 and COX-2 superimposed to illustrate similarity using UCSF Chimera.

The COX-1 enzyme is a constitutive enzyme and is present in all cells in the endoplasmic reticulum. Its concentration is maintained at a constant level. The COX-1 enzyme is a "housekeeping enzyme" involved in signaling, tissue homeostasis and cytoprotection. Most importantly, it produces prostaglandins in the stomach to help maintain the mucosal epithelium, and its inhibition causes gastric damage, bleeding and ulcers. COX-1 produced prostaglandins are required to maintain normal renal blood flow in impaired kidneys. Prostaglandins synthesized with the aid of COX-1 are essential for platelet function.

COX-2 is an enzyme induced by pain that is produced spontaneously in the range of 10–80 amplification and its concentration reduces considerably after a few hours. It is found in high concentrations around macrophages, synovial cells, mucocytes, and

fibroblasts in response to inflammation. The prostaglandins produced by COX-2 increase inflammation when the system is trying to heal the inflamed area. Figure 3 shows a comparison of the COX-1 and COX-2 enzymes.

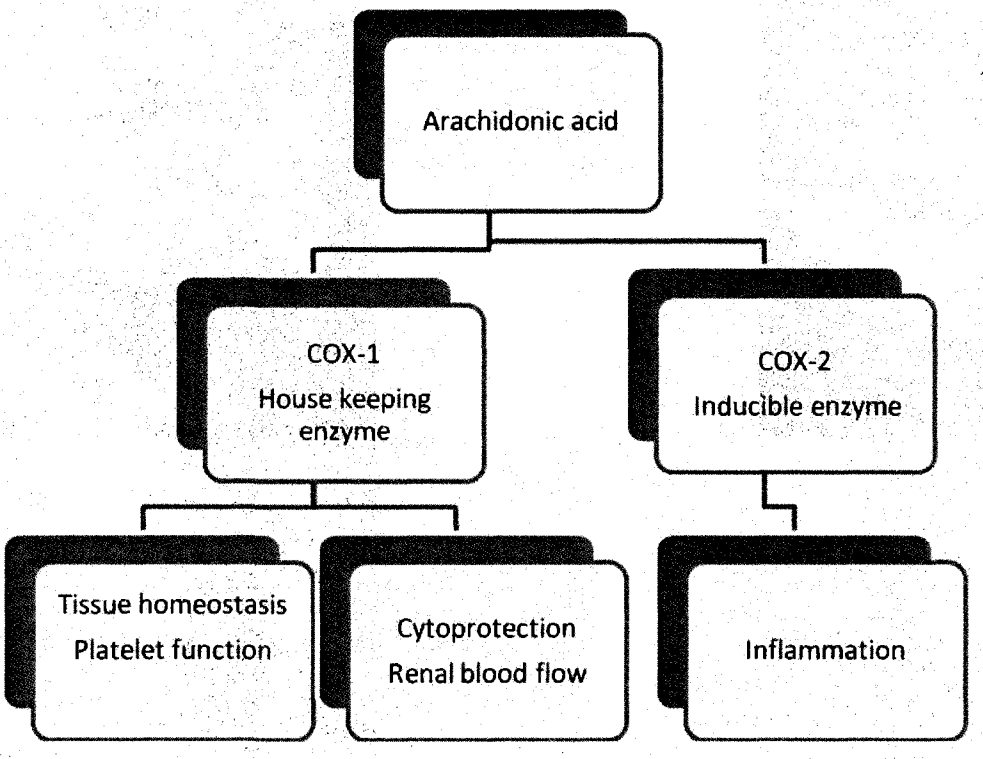


Figure 4. Comparison of COX-1 and COX-2.

The last drug target is the Interleukin-1 receptor that belongs to a family of cytokines. Cytokines are small secreted proteins which are capable of promoting inflammation. They contribute to bone degeneration in arthritis. The Interleukin-1 receptor is present in the synovial lining of the joint, and the main function of this cytokine is to cause inflammation by activating the monocytes and macrophages. Interleukin-1 stimulates prostaglandins and nitric oxide production, and promotes joint

degradation. The expression of Interleukin-1 can be stimulated by various types of cell interactions (such as TNF, autocrine or paracrine). When one type of Cytokine is inhibited the other cytokines replace it and carry its function. Other cytokines having pro-inflammatory properties can also contribute to this pathological condition, and those having anti-inflammatory properties may be able to counteract the negative effects of the former on the disease process.

1.3 Questions

1. What are the symptoms of Osteoarthritis?
2. What will happen if COX-1 mediated prostaglandin production is inhibited?
3. What will happen if COX-2 mediated prostaglandin production is inhibited?
4. Do you think it is possible to inhibit a specific class of prostaglandins? Explain.
5. What is the role of cytokines in Osteoarthritis?
6. Discuss the advantages and disadvantages of inhibiting COX-1 and COX-2.
7. Which drug target(s) would you choose for inhibition and why?
8. Perform a journal search on drug targets associated with Osteoarthritis and explain whether that drug target is inhibited.

STRUCTURE DETERMINATION MODULE-2

The objective of this module is to determine, to view and to become familiar with the target structure. The module is comprised of a set of instructions to view the structure through the VEGA ZZ program and questions and activities to know the structural details and to get acquainted with the software.

2.1 Background

Structure determination is the second step in drug design process. This step is important because the chosen drug target is a constant factor until the end of the drug design process. The drug target needs to be edited for the docking module using the VEGA ZZ program; therefore students need to understand software.

2.2 Procedure

2.2.1 Step 1 - To obtain the structure of COX-2 from Protein Data Bank

Each published 3D structure is assigned a four character identification code by Protein Data Bank. The pdb format is the most commonly used format for atomic coordinate files.The VEGA ZZ program is an interactive molecular graphics program and it needs an atomic co-ordinate file. For COX-2 structure, the pdb id is 5COX. The standard file name is 5COX.pdb.

a) In one of your windows click on the link www.pdb.org.

b) At the search engine, enter the four character protein identification code 5COX and click search.

c) Detailed information will show up on the screen about COX-2 structure.

d) Scroll down on the left hand side of the page and click on Download/Display file.

e) Click on "Complete with co-ordinates" and choose text.

f) Enter a name for the file that will contain the co-ordinates for 5COX

g) Whatever name is chosen, the file should have an extension '.pdb' for docking purpose. After entering the <name>.pdb click ok.

2.2.2 Step 2- To View the COX-2 Structure

a) Open the VEGA ZZ program.

b) Scroll down and you will find a section entitled "enter the "PDB ID". Type "5-COX" in it and click ok.

c) Now you will see the protein molecule displayed on the VEGA ZZ screen.

2.3 Questions

1. How many different components does the structure contain? (try Color by Segment)

2. Give the number of amino acids in this structure.

3. How many bound components do you see in the 5COX structure?

4. Can you locate the active site(s) in the enzyme? How?

5. What are the bound components called?

6. What is the heme group made of?

7. Find the structure of COX-1 from the PDB and display the complex in VEGA. Try to locate the ligands in the active site. Also try to read an article about this enzyme.

BINDING SITE IDENTIFICATION MODULE-3

The objective of this module is to identify the binding site in the target structure. The students are required to perform a literature search using the keywords given in the procedure.

3.1 Background

Once the target structure is determined, the binding site needs to be identified. The binding site is a place in the target structure where the ligand can bind or dock itself. The binding site in enzymes is mostly the active site. It is a hollow cavity like structure and has hydrogen bond donors and hydrophobic characteristics. Only if the ligand binds to the target and is capable of inhibiting the target, is the ligand considered a lead compound.

3.2 Step 1 - To perform a literature search to identify the binding site in the COX-2 enzyme.

a) Use the electronic database provided by the SJSU library

b) Open any scientific databases you know of like Science Direct, Wiley Interscience, etc.

c) Type the keywords "COX-2 + Active site + Location".

d) Collect and research your data.

3.3 Questions

1. What residues are around binding site?
2. There are two main active sites in the COX-2 structure. What are they? What are their functions?
3. Which binding site would you choose for your drug design module and why?
4. Point out and mark the active site in 5COX.pdb that you chose on the VEGA ZZ program.

TARGET - LIGAND DOCKING MODULE- 4

The objective of this module is to dock the ligand to the target structure using the UCSF DOCK software. The docking module is comprised of four mini modules. Each module has a set of instructions for the students to execute and answer the questions at the end. The four modules are

1. Ligand and target molecule preparation module
2. Sphere generation module
3. Scoring grid module
4. Docking module

4.1 Background

DOCK software predicts the binding modes of a small molecule-protein complex. The docking part is divided into a series of steps. Firstly, the potential sites of interest on the target molecule are identified. The site of interest includes the active site as well as other sites. This means that the software does not differentiate between the actual active site and the other potential sites.

Secondly, the software generates spheres which fill the sites of interest chosen by DOCK. The centers of the spheres are assumed to be ligand atom positions. The spheres touch the surface of the molecule but do not intersect it. The spheres are allowed to overlap or intersect other spheres. Figure 1 shows how the spheres are arranged in a sample binding site. The cyan spheres are generated with the help of points from the molecular surface (green line).

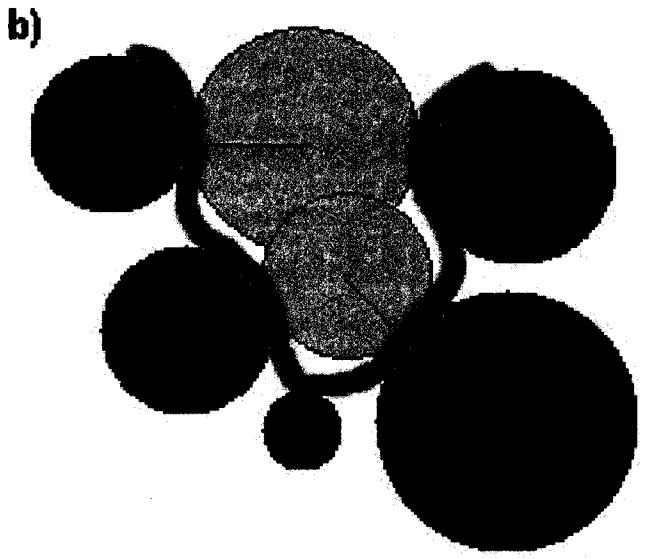


Figure 1. Representation of the spheres arranged in a binding site. [Reprinted with permission from Dr.Irwin Kuntz and Dr.Therese Lang, "Tutorial: generating spheres", http://dock.compbio.ucsf.edu/DOCK_6/tutorials/sphere_generation/generating_spheres.htm]

After the spheres are generated, the sphere centers are "matched" with the ligand atoms in order to position a ligand in the active site. Many such sphere-ligand atom sets are formed using the longest distance heuristic method. The entire orientation of the ligand atom within the active site is calculated using the sphere-ligand atom sets [10].

The orientation of the ligand is evaluated using scoring grids. Scoring grids are pre -calculated energy information for each point of the target. Every point of the target has ligand-target binding energy and an electrostatic charge. As a result, interpolated receptor values are obtained. Finally when docking is performed, the ligand values are substituted and the orientation of the ligand is evaluated based on the lowest energy score for effective binding.

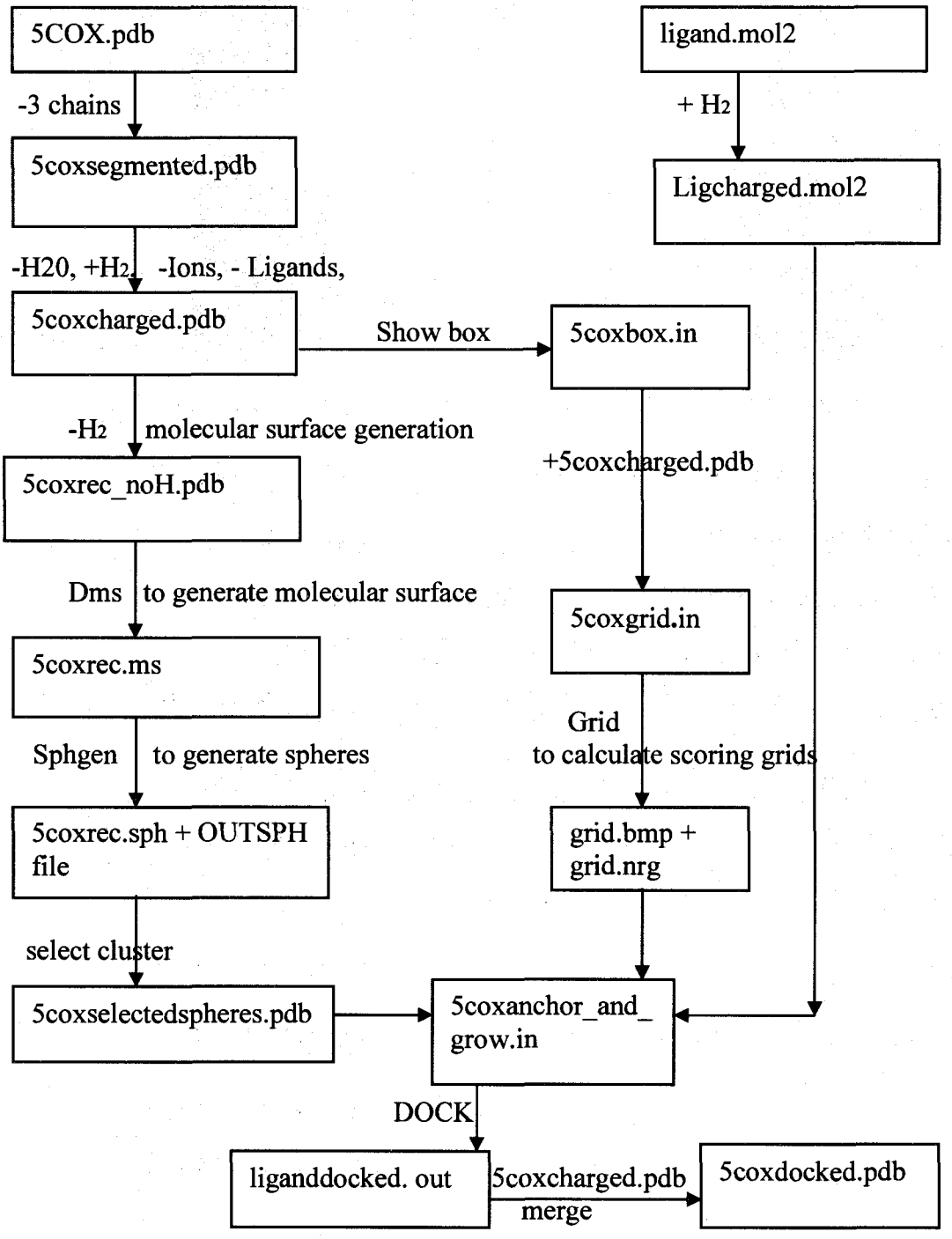
A standard method for docking described in the UCSF DOCK manual is followed. The crystal structure of COX-2 and the ligand is the starting data. To dock the target structure a step by step procedure was followed.

- Molecule preparation ligand and receptor
- Sphere generation in the receptor
- Scoring grids for the receptor
- Docking the receptor and ligand

Prior to running DOCK it is a good idea to read the following documents from these webpages

1. http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm
2. http://dock.compbio.ucsf.edu/DOCK_6/tutorials/sphere_generation/generating_spheres.htm
3. http://dock.compbio.ucsf.edu/DOCK_6/tutorials/struct_prep/prepping_molecules.htm
4. http://dock.compbio.ucsf.edu/DOCK_6/tutorials/grid_generation/generating_grid.html

The flowchart depicts the docking process in structure-based drug design.



LIGAND –RECEPTOR PREPARATION MODULE - 4.2

4.2.1 Step 1-To prepare the ligand and the receptor molecule for UCSF DOCK

Please study the content in the following webpage before starting this module;

http://dock.compbio.ucsf.edu/DOCK_6/tutorials/struct_prep/prepping_molecules.htm

The purpose of this step is to prepare the target and the ligand molecule as an input to DOCK. The drug target is usually called a receptor in the docking stage.

a) The receptor structure is a homodimer; only one subunit of COX-2 is required. Using the "Remove Segment" option in VEGA, delete all chains except one.

b) Remove the water, ions and ligand from the receptor using the "Remove" option in the VEGA ZZ program. Add hydrogen atoms with the "Add" option in the "Edit" menu.

c) Calculate the charges. Open the "Calculate" menu and click on "charges and potential". Choose Gasteiger charges to calculate.

d) Save the receptor structure as 5coxreccharged.mol2 in a separate folder called "Structure".

e) Remove hydrogen atoms from 5coxreccharged.mol2 and save it as 5coxrec_noH.pdb.

f) Obtain the ligand molecule, example LigandA from the ligand database in your system and import the mol file to the "Structure" folder.

g) Add hydrogen to the ligand, calculate Gasteiger charges, and save it as ligcharged.mol2. Verify if the valency is complete as VEGA ZZ tends to over add hydrogen atoms.

h) There are now 4 structure files, 5coxreccharged.mol2, 5coxrec_noH.pdb, ligand.mol2, and ligcharged.mol2. Work on one file at a time and always keep a backup folder.

SPHERE GENERATION MODULE – 4.3

4.3.1 Step 1 - To generate spheres in the receptor structure using Sphgen, an accessory of DOCK.

Prior to this session, it is essential to read the following webpages;

http://dock.compbio.ucsf.edu/DOCK_6/tutorials/struct_prep/prepping_molecules.

http://www.cgl.ucsf.edu/chimera/docs/UsersGuide/midas/dms1.html

a) In this session the students will use the DMS program and the Sphgen program to generate molecular surface and generate spheres on the receptor surface.

b) As a prerequisite for generating spheres in the receptor structure, the molecular surface of the receptor is generated using the Dot Molecular Surface (DMS) program.

c) The input for the dms is the 5coxrec_noH.pdb. Open the cygwin window and run the command "dms 5coxrec_noH.pdb -n –w 1.4 –v –o 5coxrec.ms" in which

- -n is used to calculate normals for surface points
- -w to change probe radius
- 1.4 is the radius
- v is the verbose
- o is the output file .

d) Save the output file as 5coxrec.ms.

e) To generate spheres over the molecular surface of the receptor, a program called sphgen is used.

f) Please note that Sphgen cannot handle more than 9999 spheres. In case of an error message, slice the receptor file to a smaller size around the periphery of the receptor structure using the VEGA ZZ program.

g) Use the "Remove atom" option in VEGA ZZ to remove atoms around the molecule.

h) Then go to step [b] and run the DMS command. Once the output file (5coxrec.ms) is obtained, proceed to step [h].

i) Create a file called INSPH in the format shown below. (Demo version of DOCK).The file contains 5coxrec.ms file, parameters, and an output file name (5coxrec.sph). Store them all in a separate folder called "Sphere-site".

File 1. The input file INSPH for sphere generation

```
5coxrec.ms
R
X
0.  0
4.  0
1.  4
5coxrec. sph
```

- R is the sphere outside of surface
- X is the subset of surface points
- 0.0 prevents generation of large spheres with close surface contacts
- 4.0 is the maximum sphere radius in Angstroms
- 1.4 is the minimum sphere radius in Angstroms

j) Open the cygwin window. Run the command "sphgen" in the same folder that contains INSPH file.

k) Two output files are generated of which one is the 5coxrec.sph file which contains spheres in clusters.The second output file is called the OUTSPH that contains calculation information. Save both files in a separate folder.

l) View the 5coxrec.sph file in the VEGA ZZ program. Then open it in pdb format and split the files according to the number of clusters. For example, if the 5coxrec.sph file has 10 clusters, put each cluster into a separate file, hence you will get 10 cluster files.

m) Merge each cluster file with the receptor file (5coxrec_noH.pdb) individually to check if the spheres in the cluster were resting on the active site using the VEGA ZZ program.

n) Based on the results of the literature search on the active site of COX-2, choose a cluster file that fills up the active site to the maximum. For example, if valine 523 is near the active site, make sure that the cluster you choose rests near that area. Save the chosen cluster file as 5coxselected spheres.sph

4.3.2 Questions

1. What is the concept behind DMS program?

2. How does Spghen generate spheres on the receptor structure?

SCORING GRID MODULE -4.4

4.4.1 Step 1 -To construct a box around the active site of the receptor structure.

Please study the content in the following webpage before starting this module; http://dock.compbio.ucsf.edu/DOCK_6/tutorials/grid_generation/generating_grid.html

Prior to grid generation, a box is constructed around the active site. The box specifies the location and size of the grid to be calculated. To create a box around the active site, a program called 'showbox' is used.

a) Generate an input file (5coxbox.in) that contains default parameters, a cluster file and the output file name in the format shown in file2 (demo version of DOCK).

b) Save the file 5coxbox.in in a separate folder called grid box.

File 2. Input file 5coxbox.in

```
y
5
5coxselectedspheres.sph
13
5coxrec_box.pdb
```

- Y- Yes (to construct box around the spheres)
- 5 - Margin to be enclosed in Angstorms.
- 13- Cluster number
- 5coxrec_box.pdb - outputfile

c) Run the command "showbox < 5coxbox.in>" and an outputfile 5coxrec_box.pdb is generated and can be view in the VEGA ZZ program.

4.4.2 Step 2 - To generate energy scoring grids for the receptor structure.

To compute energy scoring grid files, a program called Grid, an accessory of dock is used. The grid program pre-calculates the van der Waals and electrostatic interactions for the receptor active site enclosed by the grid box. The output from the calculation is stored as a scoring grid file.

$$E=\sum_{i=1}^{lig}\sum_{j=1}^{rec}\left(\frac{A_{ij}}{r_{ij}^{a}}-\frac{B_{ij}}{r_{ij}^{b}}+332\frac{q_i q_j}{Dr_{ij}}\right) \qquad \text{Equation 2}$$

- E is the intermolecular interaction energy in Kcal/mol
- i and j are the ligand and the receptor atoms
- a and b are the van der Waals repulsive and attractive exponents
- Aij and Bij are the van der Waals repulsion and attraction parameters
- r_{ij} is the distance between atoms i and j
- q_i and q_j is the point charges on the ligand i and receptor j atoms
- D is the dielectric function
- 332 is the factor to convert electrostatic energy to kcal/mol

a) Create an input file called 5coxgrid.in, consisting of the receptor file, box file and vdw definition file along with the default parameters (obtained from demo version of DOCK). File 3 shows input file 5coxgrid.in for grid program.

b) At any point if you have entered a wrong value for any parameter and would like to change it, you can edit the file and change the parameter value.

c) Run the command grid –i 5coxgrid.in. Grid generates two binary files namely grid.bmp and grid.nrg.

d) The grid.bmp file is a filter to remove ligand atoms that overlap on the receptor atoms during docking. The grid.nrg contains scoring grid information.

File 3. The input file 5coxgrid.in for grid program.

```
compute_grids                   yes
grid_spacing                    0.3
output_molecule                 no
contact_score                   no
energy_score                    yes
energy_cutoff_distance          9999
atom_model                      a
attractive_exponent             6
repulsive_exponent              12
distance_dielectric             yes
dielectric_factor               4
bump_filter                     yes
bump_overlap                    0.75
receptor_file                   5coxrec_charged.mol2
box_file                        5coxrec_box.pdb
vdw_definition_file             vdw_AMBER_parm99.defn
score_grid_prefix               grid
```

LIGAND-RECEPETOR DOCKING MODULE – 4.5

4.5.1 Step 1- To dock the ligand and the receptor using UCSF DOCK

The purpose of this step is to bind the ligand to the receptor at the site chosen (active site) in the best orientation possible. The flexible docking mode is used in this module. Flexible docking relates to the ligand and the receptor molecules as objects that change their shape spatially during the docking process.

a) Incorporate the files ligcharged.mol2, cluster file (5cox_selected spheres.sph), and the grid file (grid.nrg) into the input file called 5coxanchor_and_grow. in.

b) The default parameters from the input file of the demo version of DOCK were also added. The input file 4 shows the 5coxanchor_and _grow. in is shown as follows.

File 4. The input file 5coxanchor_and _grow.in for docking.

```
ligand_atom_file                                ligandAcharged.mol2
ligand_outfile_prefix                                         flex
limit_max_ligands                                               no
read_mol_solvation                                              no
write_orientations                                              no
write_conformations                                             no
skip_molecule                                                   no
calculate_rmsd                                                  no
rank_ligands                                                    no
num_scored_conformers_written                                    1
orient_ligand                                                  yes
automated_matching                                             yes
receptor_site_file                      5coxselectedspheres.sph
max_orientations                                               600
critical_points                                                 no
chemical_matching                                               no
use_ligand_spheres                                              no
flexible_ligand                                                yes
min_anchor_size                                                 40
num_anchor_orients_for_growth                                  100
number_confs_for_next_growth                                   100
use_internal_energy                                            yes
internal_energy_att_exp                                          6
internal_energy_rep_exp                                         12
internal energy dielectric                                     4.0
```

```
use_clash_overlap                                   no
bump_filter                                         no
score_molecules                                     yes
contact_score_primary                               no
contact_score_secondary                             no
grid_score_primary                                  yes
grid_score_secondary                                yes
grid_score_vdw_scale                                1
grid_score_es_scale                                 1
grid_score_grid_prefix                              grid
minimize_ligand                                     yes
minimize_anchor                                     yes
minimize_flexible_growth                            yes
use_advanced_simplex_parameters                     no
simplex_max_cycles                                  1
simplex_score_converge                              0.1
simplex_cycle_converge                              1.0
simplex_trans_step                                  1.0
simplex_rot_step                                    0.1
simplex_tors_step                                   10.0
simplex_anchor_max_iterations                       500
simplex_grow_max_iterations                         500
simplex_final_min_add_internal                      no
simplex_secondary_minimize_pose                     no
simplex_random_seed                                 0
atom_model                                          all
vdw_defn_file                       vdw_AMBER_parm99.defn
flex_defn_file                                 flex.defn
flex_drive_file                           flex_drive.tbl
```

c) Run the command "dock –i dock.in –i dock.out." The input file is dock.in (5coxanchor_and _grow. in) and dock.out is the output file. You can name the output file according to your choice.

d) DOCK uses all the parameters from the input file for docking, and if some data is not available, DOCK will request it from the user. You can compare your dock.in file to the dock.in file in the demo version of DOCK and use the default values as your own.

e) Please note that at any point if you have entered a wrong value for any parameter and would like to change it, you can edit the file and change the parameter value.

f) Inspect the system for few minutes to make sure that DOCK is still running and to look for errors in the input.

g) The default parameters could be changed for experimenting with the software. The output result is stored in a file called flexscored.mol2. At the end of your

output file, the lowest binding energy score of the best conformation of the ligand can be found.

h) To view the docked output in VEGA, merge the receptor file 5coxcharged.pdb with flexscored.mol2 file. Open flexscored.mol2 as pdb file in VEGA ZZ.

i) To dock other ligands B, C, and D with the receptor follow the modules 4.2 (g, h) and then 4.5.

j) The ligands might generate an error message that it could not complete growth and it might suggest to increase the size of the grid box and to increase the ligand orientations. Try increasing the grid box and the maximum orientations in the input file. If the error message repeats, it could be because the binding site is too small or the ligand molecule is too large to fit inside the grid box and the active site.

POTENTIAL LEAD EVALUATION MODULE- 5

This module aims to evaluate the ligands that were docked based on Lipinski's Rule of 5 and the number of rotatable bonds. It also explains the criterion for evaluating the ligands based on their binding affinity obtained from DOCK. Questions at the end of this module will help the students analyze ligands and come up with a potential lead.

5.1 Background

Once the ligand has been successfully identified, it is called the lead compound. It must be evaluated before proceeding to the preclinical trials. The lead is generally evaluated for its bioavailability, oral viability, chemical and physical stability, and ease of production. The final output file has a binding or grid score expressed as a sum of van der Waals forces and electrostatic interactions. The more negative the score is, the higher the binding affinity. Based on Veber's paper, the number of rotatable bonds should be less than 10 for oral bioavailability. Lipinski's Rule of 5 is followed for the lead drug evaluation. It is called Lipinski's Rule of 5 because all the four parameters are close to 5 or multiples of 5.

Lipinski's rule states that poor absorption or permeation is more likely when

- There are more than five H-bond donors
- The molecular weight is over 500 Da
- The logP (partition coefficient) is over 5
- The sums of N's and O's are over 10

5.2 Step 1- To evaluate the docked ligand.

Please read the journal article "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings" by Lipinski et al.

a) The chemical names of ligands will be disclosed. Tabulate the characteristics of the docked ligands based on the Lipinski's Rule of 5, the number of rotatable bond and the binding scores obtained from DOCK.

b) Plot a partition co-efficient vs molecular weight graph from your table.

c) Analyze your ligand and come up with the best ligand.

5.3 Questions

a) State the Lipinski's rule of 5.

b) Why is a Log P greater than 5 unfavorable to the ligand?

c) A lead drug for HIV infected patients is out in the market. It does not obey the Lipinski's rules. What are the disadvantages of that drug?

APPENDIX C

INFORMATION AND ANSWER KEY FOR INSTRUCTORS

DRUG TARGET IDENTIFICATION MODULE-1

A) Choosing the Right Drug Target

Analyzing the given drug targets, it is apparent that prostaglandins which are responsible for pain cannot be inhibited because they are produced by two different cyclooxygenases which carry out different functions altogether. The prostaglandins produced by COX-1 are essential for cytoprotection in the stomach, platelet aggregation and renal function. Prostaglandins produced by COX-2 cause inflammation. If inhibition of prostaglandins would not only stop inflammation but also stop positive functions of the body and cause damaging side effects. Hence prostaglandins cannot be used as a drug target.

As for cytokines, they cannot be used as a drug target because specific blockage of the Interleukin-1 receptor causes other cytokines to replace its function. Therefore, it would not be a worthwhile strategy to inhibit the Interleukin-1 receptor.

It is evident that the COX-1 enzyme also cannot be used as a drug target because inhibiting this cyclooxygenase isoform causes gastric damage, hemorrhageing and ulceration in the stomach. Inhibition of COX-1 also causes irregular blood flow in the kidney and platelet dysfunction in the circulatory system. The presence of COX-1 in all cells is important for cell function such as homeostasis.

Although COX-1 and COX-2 enzymes are 60% homologous in structure they carry out different functions; one maintains normal cellular physiology while the other

causes pain. COX-2 enzyme is upregulated during inflammation and is present in insignificant amounts at other times in the human system; hence its inhibition would not affect the normal functions of the body. Since the aim of treatment is pain control with few side effects, COX-2 enzyme is the ideal drug target. Inhibition of the COX-2 enzyme would result in anti-inflammatory and analgesic effects.

B) Answer key for Drug Target Identification Module

1. What are the symptoms of Osteoarthritis?

Answer: Inflammation, tissue irritation and pain in the joints are the s symptoms of Osteoarthritis.

2. What will happen if we inhibit prostaglandins that are produced by COX-1?

Answer: COX-1 produced prostaglandins are required to maintain normal renal blood flow in impaired kidneys. The prostaglandins are essential for cytoprotection in the stomach, platelet aggregation and renal function.

3. What will happen if we inhibit prostaglandins that are produced by COX-2?

Answer: Prostaglandins produced by COX-2 cause inflammation.

4. Do you think it is possible to inhibit a specific kind of prostaglandin? Explain.

Answer: Prostaglandins which are responsible for pain, cannot be inhibited because they are produced by two different cyclooxygenases which carry out different functions altogether. The prostaglandins produced by COX-1 are essential for cytoprotection in the stomach, platelet aggregation and renal function. Prostaglandins produced by COX-2 cause inflammation. If prostaglandins as a class are inhibited it would not only stop

inflammation but also stop normal cellular functions of the body and cause damaging side effects. Hence prostaglandins cannot be inhibited.

5. What is the role of cytokines in Osteoarthritis?

Answer: Cytokines are small secreted proteins which are pro-inflammatory. They contribute to the bone degeneration of arthritis. The Interleukin-1 receptor is present in the synovial lining of the joint. The main function of this particular cytokine is to cause inflammation by activating the monocytes and macrophages.

6. Discuss disadvantages of inhibiting COX-1 and advantages of inhibiting COX-2.

Answer: The COX-1 enzyme on inhibition causes gastric damage, hemorrhaging and ulceration in the stomach. The inhibition also causes irregular blood flow in the kidney and platelet dysfunction in the circulatory system. COX-2 enzyme is upregulated during inflammation and is present at other times in insignificant amounts in the human system; hence its inhibition would not affect the normal functions of the body. Inhibition of the COX-2 enzyme would result in anti-inflammatory and analgesic effects.

7. Which drug target(s) is the root cause of inflammation and how?

Answer: The COX-2 enzyme is an induced enzyme that is increased in concentration in the range of 10–80 in response to injury (its count reduces considerably after a few hours). It is found in high concentrations around macrophages, synovial cells, mucocytes, and fibroblasts in response to inflammation. COX-2 enzyme is characteristic of inflammation when the system is trying to heal the inflamed area.

STRUCTURE DETERMINATION MODULE-2

A) Installation Requirements

Install VEGA ZZ software for this module from the website:

http://www.ddl.unimi.it/vega/download.htm

B) Answer key to the Structure Determination Module

1. How many different components does the structure contain?

Answer: The COX-2 structure is a homodimer of identical subunits. The COX-2 structure comprises of four protein chains.

2. Give the number of aminoacids in this structure.

Answer: 587 amino acids

3. How many bound components do you see in the 5COX structure?

Answer: The structure contains four heme ligands and 12 N-acetyl glucosamine (NAG) trisaccharide ligands.

4. Can you locate the active site(s) in the enzyme? How?

Answer: The cyclooxygenase active site is located near valine 523.

5. What is the heme group made of?

Answer: The heme group is protoporphyrin IX containing iron (Fe). The metal free porphyrin combines with the ferrous ion to form the heme group.

BINDING SITE IDENTIFICATION MODULE-3

A) Answer key Binding Site Identification Module

1. What is an active site?

Answer: The specific part at which the substarte attaches to the enzyme.

2. What is the difference between active and binding site?

Answer: They are the same more or less.

3. What residues are around binding site?

Answer: Valine 523, Phenyalanine 518, Valine 434, Tyrosine 385, Arginine 513.

4. What are the two main active sites in the COX-2 structure? What are their functions?

Answer: The COX-2 enzyme structure has two different active sites, the cyclooxygenase and the peroxidase active site. At the cyclooxygenase active site, arachidonate binds at the site to form prostaglandins. Peroxidase active site is needed to reduce a peroxy intermediate to an alcohol and to activate the heme groups that participate in the cyclooxygenase reaction. The cyclooxygenase active site is chosen to be the correct active site and will be used for the docking stage.

5. Which binding site would you choose for your drug design module and why?

Answer: The cyclooxygenase active site is located near valine 523, phenyalanine 518, valine 434, and tyrosine 385. Since the valine molecule is small, it leaves a gap causing a side pocket. This side pocket is the site of binding for many selective drugs.

6. Point out and mark the active site in 5COX.pdb that you chose on the VEGA ZZ program.

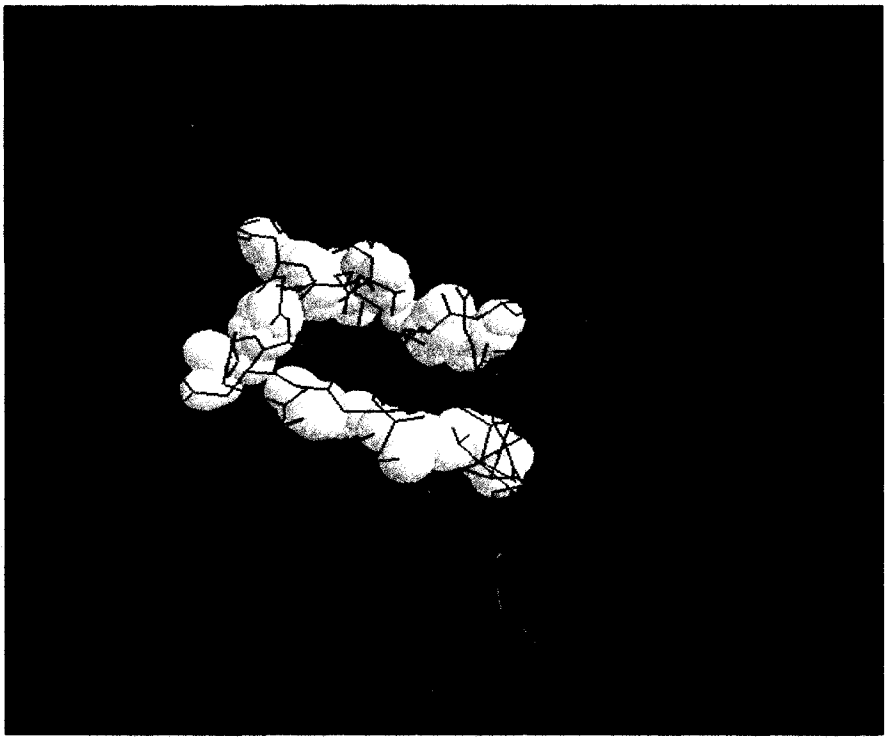


Figure 2. COX-2 structure with active site surrounded by white residues.

TARGET - LIGAND DOCKING MODULE- 4

A. Installation Requirements for Target–Ligand Docking module

a. To obtain Academic Software License for UCSF DOCK please access http://dock.compbio.ucsf.edu/Online_Licensing/dock_license_application.html

b. After receiving the license for UCSF DOCK, you will receive the link to install the software. Follow the instructions based on the system configuration that SJSU provides and install the software.

c. A copy of the DOCK reference manual is necessary to the instructors.

B. The Ligands' Information

The ligands' names are not disclosed to the students. The information about potential ligand compounds are obtained from KEGG Ligand Database. The chemical names and mol formats for four ligands are as follows. Please store them as Ligand A, B, C, and D in a folder called the ligand database. The actual names are as follows

f. Ligand A = Acetyl salicylic acid

g. Ligand B = Rofecoxib

h. Ligand C = Celecoxib

i. Ligand D = SC-558 (1-PHENYLSULFONAMIDE-3-TRIFLUOROMETHYL-5-PARABROMOPHENYLPYRAZOLE)

j. Ligand E = Fucoxanthin

KEGG COMPOUND: C01405

Help

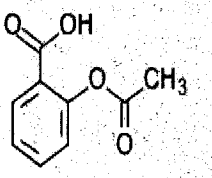
Entry	C01405 Compound
Name	Aspirin; Acetylsalicylic acid; 2-Acetoxybenzenecarboxylic acid; Acetylsalicylate
Formula	C9H8O4
Mass	180.0423
Structure	 C01405 Mol file KCF file DB search
Remark	Same as: D00109
Reaction	R02942
Enzyme	1.14.99.1 (I) 3.1.1.55
Other DBs	CAS: 50-78-2 PubChem: 4594 ChEBI: 15365 3DMET: B00284
LinkDB	All DBs
KCF data	Show

Figure 1. Information for Ligand A in the KEGG Ligand Database.

File 1. The mol file for Ligand A.

```
  13 13  0  0  0  0  0  0  0  0999 V2000
   20.2981  -15.8105    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   21.5226  -16.5029    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   19.0928  -16.5029    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   20.2981  -14.6927    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   21.5226  -17.9133    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   22.7278  -15.8040    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   19.0928  -17.9133    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   21.5033  -13.9940    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   19.0863  -14.0004    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   20.2981  -18.6250    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   23.9396  -16.4964    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   25.1450  -15.7977    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   23.9396  -17.9642    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0     0  0
  1  3  2  0     0  0
  1  4  1  0     0  0
  2  5  2  0     0  0
  2  6  1  0     0  0
  3  7  1  0     0  0
  4  8  1  0     0  0
  4  9  2  0     0  0
  5 10  1  0     0  0
  6 11  1  0     0  0
 11 12  1  0     0  0
 11 13  2  0     0  0
  7 10  2  0     0  0
M  END
```

KEGG **COMPOUND: C07590**

Help

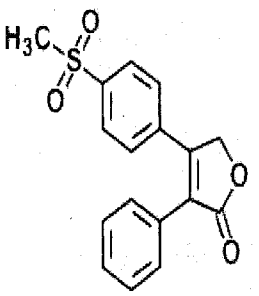
Entry	C07590 Compound
Name	Rofecoxib
Formula	C17H14O4S
Mass	314.0613
Structure	 C07590 Mol file KCF file DB search
Remark	Same as: D00568
Other DBs	CAS: 162011-90-7 PubChem: 9792 ChEBI: 8887
LinkDB	All DBs
KCF data	Show

Figure 2. Information of Ligand B in the KEGG Ligand Database.

File 2. Mol file for Ligand B exported from KEGG Ligand Database.

```
22 24  0  0  0  0  0  0  0  0999 V2000
   29.9147  -19.5829    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   29.0770  -18.4195    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   27.7739  -18.8384    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   27.7739  -20.2810    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   29.0770  -20.6999    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   29.5424  -22.0030    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   24.1439  -20.9791    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   24.1439  -22.3753    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   25.3539  -23.0734    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   26.5638  -22.3753    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   26.5638  -20.9791    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   25.3539  -20.2810    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   24.1439  -16.7441    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   24.1439  -18.1403    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   25.3539  -18.8384    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   26.5638  -18.1403    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   26.5638  -16.7441    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   25.3539  -16.0460    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   22.9338  -16.0460    0.0000 S   0  0  0  0  0  0  0  0  0  0  0  0
   23.6319  -14.8359    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   22.1892  -17.2560    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   21.6772  -15.3479    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0     0  0
  2  3  1  0     0  0
  3  4  2  0     0  0
  4  5  1  0     0  0
  1  5  1  0     0  0
  5  6  2  0     0  0
  7  8  2  0     0  0
  8  9  1  0     0  0
  9 10  2  0     0  0
 10 11  1  0     0  0
 11 12  2  0     0  0
  7 12  1  0     0  0
 11  4  1  0     0  0
 13 14  2  0     0  0
 14 15  1  0     0  0
 15 16  2  0     0  0
 16 17  1  0     0  0
 17 18  2  0     0  0
 13 18  1  0     0  0
 16  3  1  0     0  0
 13 19  1  0     0  0
 19 20  2  0     0  0
 19 21  2  0     0  0
 19 22  1  0     0  0
M  END
```

KEGG DRUG: D00567

Help

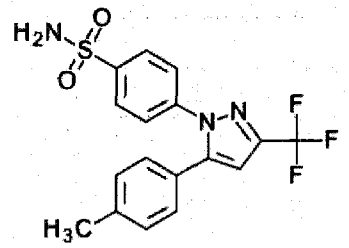
Entry	D00567 Drug
Name	Celecoxib (JAN/USAN/INN); Celebrex (TN)
Formula	C17H14F3N3O2S
Mass	381.0759
Structure	 D00567 Mol file KCF file DB search
Target	cyclooxygenase-2 (COX-2) inhibitor [HSA:5743] [EC:1.14.99.1]
Activity	Anti-inflammatory and analgesic [cyclooxygenase-2 inhibitor]
Remark	Same as: C07589 Therapeutic category: 1149
Pathway	PATH: map07112 1,2-Diphenyl substitution family
Other DBs	CAS: 169590-42-5 PubChem: 7847633
LinkDB	All DBs
KCF data	Show

Figure 3. Information of Ligand C in the KEGG Ligand Database.

File 3. Mol file for Ligand C exported from KEGG Ligand Database.

```
 26 28  0  0  0  0  0  0  0  0999 V2000
    23.9641  -17.9834    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    25.3657  -17.9834    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    23.1231  -16.8622    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    21.8150  -17.2826    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    21.8150  -18.6842    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    23.1231  -19.1047    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    25.3657  -19.3850    0.0000 F   0  0  0  0  0  0  0  0  0  0  0  0
    26.7672  -17.9834    0.0000 F   0  0  0  0  0  0  0  0  0  0  0  0
    25.3657  -16.5819    0.0000 F   0  0  0  0  0  0  0  0  0  0  0  0
    18.1709  -15.1803    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    18.1709  -16.5819    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    19.3856  -17.2826    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    20.6003  -16.5819    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    20.6003  -15.1803    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    19.3856  -14.4795    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    18.1709  -19.3850    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    18.1709  -20.7866    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    19.3856  -21.4874    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    20.6003  -20.7866    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    20.6003  -19.3850    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    19.3856  -18.6842    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    16.9572  -21.4876    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    16.9572  -14.4793    0.0000 S   0  0  0  0  0  0  0  0  0  0  0  0
    17.6570  -13.2647    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    16.2554  -15.6942    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    15.7415  -13.7787    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0     0  0
  1  3  2  0     0  0
  3  4  1  0     0  0
  4  5  1  0     0  0
  5  6  2  0     0  0
  1  6  1  0     0  0
  2  7  1  0     0  0
  2  8  1  0     0  0
  2  9  1  0     0  0
 10 15  1  0     0  0
 13  4  1  0     0  0
 16 17  2  0     0  0
 17 18  1  0     0  0
 18 19  2  0     0  0
 19 20  1  0     0  0
 20 21  2  0     0  0
 16 21  1  0     0  0
 20  5  1  0     0  0
 17 22  1  0     0  0
 10 23  1  0     0  0
 23 24  2  0     0  0
 23 25  2  0     0  0
 23 26  1  0     0  0  M  END
```

File 4. Ligand D pdb file exported from KEGG Ligand Database.

```
ATOM      1  C1   S58  6380       1.383    0.005   -2.987  1.00 20.00    C+0
ATOM      2  C2   S58  6380       0.862    0.013   -1.724  1.00 20.00    C+0
ATOM      3  C3   S58  6380       0.319   -0.006   -3.890  1.00 20.00    C+0
ATOM      4  C4   S58  6380       0.440   -0.014   -5.392  1.00 20.00    C+0
ATOM      5  C5   S58  6380      -1.422   -0.002   -0.794  1.00 20.00    C+0
ATOM      6  C6   S58  6380      -2.578    0.762   -0.883  1.00 20.00    C+0
ATOM      7  C7   S58  6380      -3.488    0.759    0.155  1.00 20.00    C+0
ATOM      8  C8   S58  6380      -3.249   -0.004    1.283  1.00 20.00    C+0
ATOM      9  C9   S58  6380      -2.096   -0.762    1.377  1.00 20.00    C+0
ATOM     10  C10  S58  6380      -1.185   -0.767    0.339  1.00 20.00    C+0
ATOM     11  C11  S58  6380       1.628    0.028   -0.459  1.00 20.00    C+0
ATOM     12  C12  S58  6380       2.785   -0.741   -0.332  1.00 20.00    C+0
ATOM     13  C13  S58  6380       3.496   -0.724    0.849  1.00 20.00    C+0
ATOM     14  C14  S58  6380       3.063    0.054    1.909  1.00 20.00    C+0
ATOM     15  C15  S58  6380       1.916    0.820    1.789  1.00 20.00    C+0
ATOM     16  C16  S58  6380       1.200    0.815    0.610  1.00 20.00    C+0
ATOM     17  N1   S58  6380      -0.498   -0.001   -1.847  1.00 20.00    N+0
ATOM     18  N2   S58  6380      -0.802   -0.010   -3.214  1.00 20.00    N+0
ATOM     19  N3   S58  6380      -5.494   -1.224    2.313  1.00 20.00    N+0
ATOM     20  O1   S58  6380      -5.148    1.203    2.475  1.00 20.00    O+0
ATOM     21  O2   S58  6380      -3.685   -0.404    3.760  1.00 20.00    O+0
ATOM     22 BR1   S58  6380       4.043    0.072    3.526  1.00 20.00   BR+0
ATOM     23  F1   S58  6380      -0.836   -0.026   -5.963  1.00 20.00    F+0
ATOM     24  F2   S58  6380       1.127    1.130   -5.810  1.00 20.00    F+0
ATOM     25  F3   S58  6380       1.143   -1.154   -5.797  1.00 20.00    F+0
ATOM     26  S1   S58  6380      -4.413   -0.005    2.606  1.00 20.00    S+0
END
```

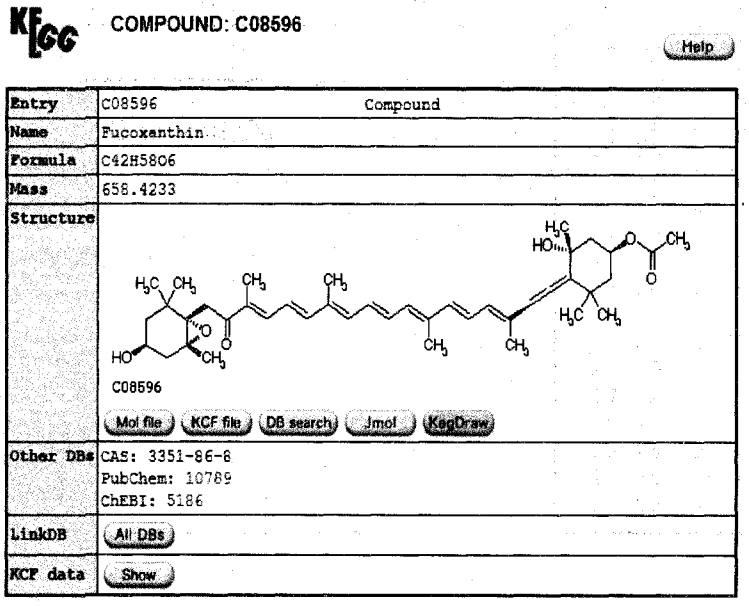


Figure 4. Information of Ligand E in the KEGG Ligand Database.

File 14. The mol file for Ligand E.

```
ISISHOST03240423202D 1   1.00000     0.00000  7636

 48 50  0     1  0            999 V2000
   -4.6414   -0.3483    0.0000 C   0  0  1  0  0  0            0  0  0
   -4.6414   -0.8759    0.0000 C   0  0  2  0  0  0            0  0  0
   -4.1931   -0.6103    0.0000 O   0  0  0  0  0  0            0  0  0
   -5.1000   -0.0828    0.0000 C   0  0  3  0  0  0            0  0  0
   -4.1931   -0.0828    0.0000 C   0  0  0  0  0  0            0  0  0
   -5.1000   -1.1379    0.0000 C   0  0  0  0  0  0            0  0  0
   -4.1931   -1.1379    0.0000 C   0  0  0  0  0  0            0  0  0
   -5.5655   -0.3483    0.0000 C   0  0  0  0  0  0            0  0  0
   -5.3690    0.3724    0.0000 C   0  0  0  0  0  0            0  0  0
   -4.8414    0.3724    0.0000 C   0  0  0  0  0  0            0  0  0
   -3.7345   -0.3448    0.0000 C   0  0  0  0  0  0            0  0  0
   -5.5655   -0.8759    0.0000 C   0  0  2  0  0  0            0  0  0
   -3.2828   -0.0793    0.0000 C   0  0  0  0  0  0            0  0  0
   -3.7414   -0.8690    0.0000 O   0  0  0  0  0  0            0  0  0
   -6.0000   -1.1345    0.0000 O   0  0  0  0  0  0            0  0  0
   -2.8276   -0.3414    0.0000 C   0  0  0  0  0  0            0  0  0
   -3.2897    0.4483    0.0000 C   0  0  0  0  0  0            0  0  0
   -2.3724   -0.0759    0.0000 C   0  0  0  0  0  0            0  0  0
   -1.9172   -0.3379    0.0000 C   0  0  0  0  0  0            0  0  0
   -1.4621   -0.0724    0.0000 C   0  0  0  0  0  0            0  0  0
   -1.0069   -0.3345    0.0000 C   0  0  0  0  0  0            0  0  0
   -1.4655    0.4517    0.0000 C   0  0  0  0  0  0            0  0  0
   -0.5517   -0.0724    0.0000 C   0  0  0  0  0  0            0  0  0
   -0.0966   -0.3345    0.0000 C   0  0  0  0  0  0            0  0  0
    0.3552   -0.0690    0.0000 C   0  0  0  0  0  0            0  0  0
    0.8103   -0.3310    0.0000 C   0  0  0  0  0  0            0  0  0
    1.2621   -0.0655    0.0000 C   0  0  0  0  0  0            0  0  0
    0.8069   -0.8586    0.0000 C   0  0  0  0  0  0            0  0  0
    1.7207   -0.3276    0.0000 C   0  0  0  0  0  0            0  0  0
    2.1724   -0.0586    0.0000 C   0  0  0  0  0  0            0  0  0
    2.6276   -0.3241    0.0000 C   0  0  0  0  0  0            0  0  0
    3.0862   -0.0586    0.0000 C   0  0  0  0  0  0            0  0  0
    2.6241   -0.8517    0.0000 C   0  0  0  0  0  0            0  0  0
    3.5379    0.2069    0.0000 C   0  0  0  0  0  0            0  0  0
    3.9897    0.4724    0.0000 C   0  0  0  0  0  0            0  0  0
    4.4448    0.2103    0.0000 C   0  0  3  0  0  0            0  0  0
    3.9897    1.0000    0.0000 C   0  0  2  0  0  0            0  0  0
    4.9034    0.4724    0.0000 C   0  0  0  0  0  0            0  0  0
    4.7069   -0.2448    0.0000 C   0  0  0  0  0  0            0  0  0
    4.1793   -0.2448    0.0000 C   0  0  0  0  0  0            0  0  0
    4.4448    1.2621    0.0000 C   0  0  0  0  0  0            0  0  0
    3.8517    1.5103    0.0000 C   0  0  0  0  0  0            0  0  0
    3.4793    1.1379    0.0000 O   0  0  0  0  0  0            0  0  0
    4.9034    1.0000    0.0000 C   0  0  1  0  0  0            0  0  0
    5.3621    1.2655    0.0000 O   0  0  0  0  0  0            0  0  0
    5.8138    1.0035    0.0000 C   0  0  0  0  0  0            0  0  0
    6.2655    1.2655    0.0000 C   0  0  0  0  0  0            0  0  0
    5.8069    0.4759    0.0000 O   0  0  0  0  0  0            0  0  0
  1  2  1  0     0  0
  1  3  1  6     0  0
  1  4  1  0     0  0
  1  5  1  1     0  0
  2  6  1  0     0  0
  2  7  1  1     0  0
  4  8  1  0     0  0
```

SPHERE GENERATION MODULE – 4.3

A) Installation Requirements for Sphere generation module

Please install the Dot Molecular Surface program (DMS) from the website http://www.cgl.ucsf.edu/chimera/docs/UsersGuide/midas/dms1.html. The procedure for DMS was followed from the "Tutorial Generating Spheres" website which is a part of the official DOCK website; http://dock.compbio.ucsf.edu/DOCK_6/tutorials/sphere_generation/generating_spheres.

This website gives clear instructions regarding molecular surface generation and sphere genearation. The output file (5coxrec.sph file) obtained from sphgen was edited by splitting into 49 different cluster files. The 5coxrec.sph file is shown in pdb format in Figure 1.

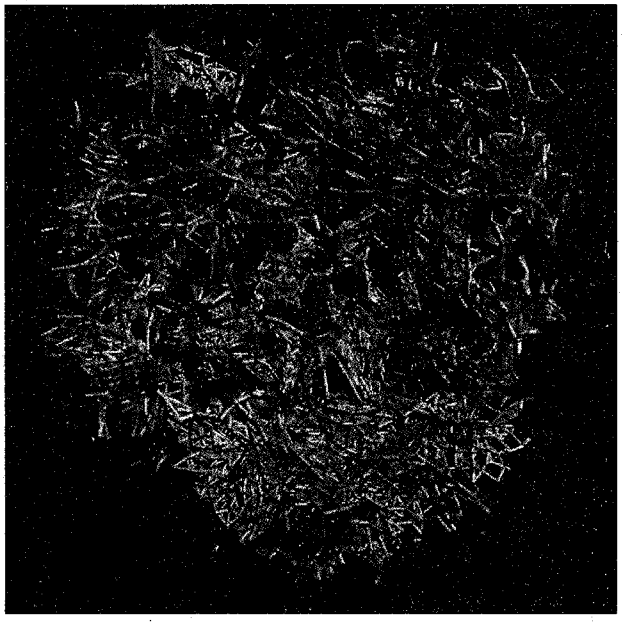


Figure 1. The 5coxrec.sph file is shown in pdb format.

Then the cluster files are merged with the modified receptor file (5coxrec_noH.pdb) individually. The VEGA ZZ program is used to check if the spheres in the cluster were resting on the active site. It is found that the 5th cluster file matches with the active site of the receptor.

Please give fifth cluster file shown in File 1 to students as they might have chosen a different cluster file. The cluster file should be saved as 5coxselectedspheres.sph.

File 1. The fifth cluster file.

```
cluster     5   number of spheres in cluster    28
  564   25.47775   21.24536   18.13414    2.069 1440 0   0
 1407   24.89094   23.01594   15.34170    2.629 2382 0   0
 1408   25.53015   23.88636   14.41222    2.687 2379 0   0
 1432   24.32610   24.15849   16.15387    1.964 1407 0   0
 1433   24.53161   23.30254   14.48500    2.882 2356 0   0
 1435   24.48589   23.48287   13.64364    2.930 2356 0   0
 1440   24.55522   21.06678   17.15050    2.280  564 0   0
 1455   26.13715   21.23123   19.92801    1.536 1494 0   0
 1456   24.61458   20.51658   17.57876    1.950  564 0   0
 1494   26.41505   21.02241   19.82538    1.653 1455 0   0
 1713   25.27725   24.35810   10.50218    1.669 1724 0   0
 1724   25.48431   24.41288   11.92778    2.226 2376 0   0
 1748   24.78060   23.99746   12.08374    2.347 2375 0   0
 1750   24.86185   24.11027   12.70872    2.652 2376 0   0
 2320   22.65622   23.47803   14.64984    1.740 1431 0   0
 2345   23.89269   23.47197   11.35615    2.015 1748 0   0
 2351   24.74022   22.05213   13.03681    1.837 2378 0   0
 2355   24.42115   21.02557   17.17635    2.223 1440 0   0
 2356   24.56858   23.41748   13.93200    2.932 2379 0   0
 2375   24.12689   23.52305   11.48329    2.140 1748 0   0
 2376   24.56398   23.84492   12.97488    2.786 1748 0   0
 2378   24.70856   21.80117   12.29410    1.569 2351 0   0
 2379   24.87683   23.63687   14.30022    2.894 1435 0   0
 2382   24.49579   22.54498   15.31068    2.528 2356 0   0
 2405   27.95099   23.52323   15.73837    1.740 1408 0   0
 2412   28.46263   22.79210   15.92203    1.688 2379 0   0
 2413   27.56845   21.04100   18.21653    1.837  564 0   0
 2414   29.08758   23.93856   16.37423    1.428 1408 0   0
```

Figure 2 shows the spheres of the cluster file merging with the active site of the receptor file. The fifth cluster contains 28 spheres shown as solid spheres. The tubular structures are valine 523 (yellow) and phenylalanine 518 (purple) atoms which are near the active site.

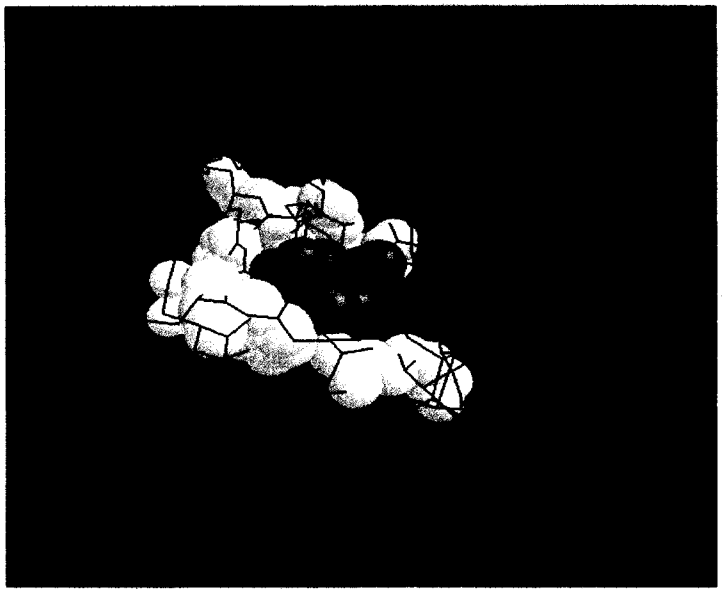


Figure 2. Spheres of the cluster file resting on the active site

B) Answer key to the sphere generation module

1. What is the concept behind DMS program?

Answer: The DMS program generates molecular surface based on an algorithm developed by Richards and Conolly. It is viewed as van der Waals surface of a molecule except that the crevices are softened and the cavities are included.

2. How does Spghen generate spheres on the receptor structure?

Answer: The software generates spheres which fill the sites of interest chosen by DOCK. The centers of the spheres are assumed to be ligand atom positions. The spheres touch the surface of the molecule but do not intersect it. The spheres are allowed to overlap or intersect other spheres. After the spheres are generated, the sphere centers are "matched" with the ligand atoms in order to position a ligand in the active site. Many such sphere-ligand atom sets are formed using the longest distance heuristic method.

LIGAND–RECEPTOR DOCKING MODULE

A) Table 1. The binding scores of the docked ligand.

Ligand	Grid score for flexible docking Kcal/mol	Grid score for rigid docking Kcal/mol
Ligand A(acetyl salicylic acid)	-38.7330	-36.6905
Ligand B(rofecoxib)	-40.356	-35.400
Ligand C(celecoxib)	-34.592	-14.554
Ligand D(SC-558)	- 40.7028	13.4299
Ligand E(fucoxanthin)	Error: Could not complete growth. Confirm grid box is large enough to contain ligand	Error: Conformation could not be scored by DOCK. Conformation not completely within grid box.

POTENTIAL LEAD EVALUATION MODULE- 5

A) Table shows the Lipinski's rule of 5 characteristics for the docked ligands.

Ligand	Xlogp Partition coefficient	OH+NH H-bond donors	Mol Wt g/mol	N+O H-bond acceptors	Rotatable bonds	Alert 1: good; 0: poor
Ligand A	1.4	1	180.16	4	3	1
Ligand B	3.2	0	314.357	4	3	1
Ligand C	3.9	1	381.373	8	3	1
Ligand D	4.2	1	446.243	8	3	1
Ligand E(fucoxanthin)	7.9	2	658.906	6	12	0

B) Answer key to potential lead evaluation module.

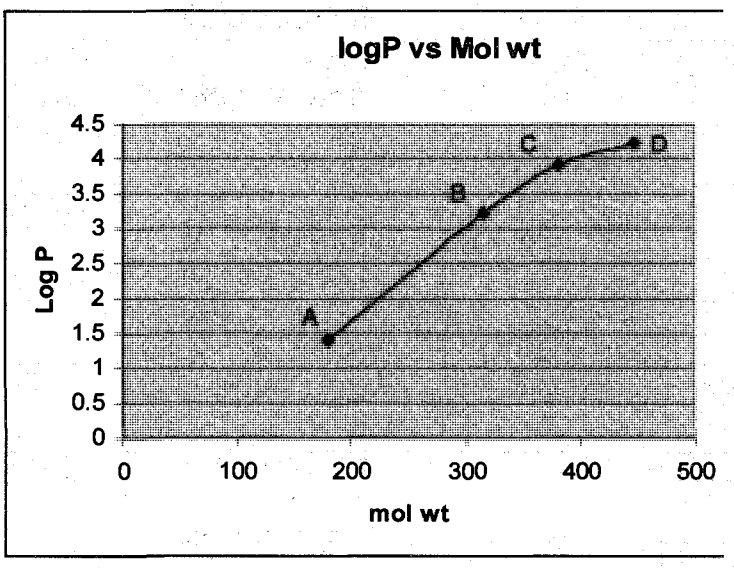
1. State the Lipinski's rule of 5.

Answer: Lipinski's rule states that poor absorption or permeation is more likely when

- There are more than five H-bond donors
- The molecular weight is over 500 Da
- The logP (partition coefficient) is over 5
- The sums of N's and O's are over 10

2. Why is a Log P greater than 5 unfavorable to the ligand?

Answer: It is known that permeability decreases with high molecular weight, resulting in less bio absorption. MLog P is the partition coefficient and is found to be directly proportional to the molecular weight. As the molecular weight increases, the log P values become larger, indicating less permeability.



Graph showing logP vs. molecular weight

3. A lead drug for HIV infected patients is out in the market. It does not obey the Lipinski's rules. What are the disadvantages of that drug?

Answer: If a lead drug does not obey the Lipinski's rule of 5, it means that the drug would not be absorbed properly. Even if they are absorbed, it may not survive when it enters the bloodstream. As HIV patients have weak immune systems it is not a good idea to give the drug to them, hence it is not a good drug.